

Nos. 14-1361, -1366

**United States Court of Appeals
For The Federal Circuit**

IN RE BRCA1- AND BRCA2-BASED HEREDITARY
CANCER TEST PATENT LITIGATION

UNIVERSITY OF UTAH RESEARCH FOUNDATION,
THE TRUSTEES OF THE UNIVERSITY OF PENNSYLVANIA,
HSC RESEARCH AND DEVELOPMENT LIMITED PARTNERSHIP,
ENDORECHERCHE, INC., AND MYRIAD GENETICS, INC.,

Plaintiffs-Appellants,

v.

AMBRY GENETICS CORPORATION,

Defendant-Appellee.

Appeals from the United States District Court for the Central District of Utah
in consolidated case Nos. 2:13-cv-00640-RJS and 2:14-md-02510-RJS,
Judge Robert J. Shelby.

APPELLANTS' OPENING BRIEF

David G. Mangum
Michael R. McCarthy
Kristine E. Johnson
C. Kevin Speirs
PARSONS BEHLE & LATIMER
201 South Main Street, Suite 1800
P.O. Box 45898
Salt Lake City, UT 84111
(801) 532-1234

*Attorney for Appellants
University of Utah Research Foundation, the
Trustees of the University of Pennsylvania,
HSC Research and Development Limited
Partnership, Endorecherche, Inc., and
Myriad Genetics, Inc.*

Jonathan E. Singer
Deanna J. Reichel
FISH & RICHARDSON P.C.
3200 RBC Plaza
60 South Sixth Street
Minneapolis, MN 55402
(612) 335-5070

Geoff D. Biegler
FISH & RICHARDSON P.C.
12390 El Camino Real
San Diego, CA 92130
(858) 678-5070

*Attorneys for Appellant
Myriad Genetics, Inc.*

UNIVERSITY OF UTAH, et al. v. AMBRY GENETICS

14-1361, -1366

CERTIFICATE OF INTEREST

Counsel for the Plaintiff-Appellant, Myriad Genetics, Inc. certifies the following:

1. The full name of every part or amicus represented by me is:

Myriad Genetics, Inc.

2. The name of the real party in interest represented by me is:

Myriad Genetics, Inc.

3. All parent corporations and any publicly held companies that own 10 percent or more of the stock of the party or amicus curiae represented by me are:

None.

4. The names of all law firms and the partners or associates that appeared for the party or amicus now represented by me in the trial court or agency or are expected to appear in this court are:

Fish & Richardson P.C.: Jonathan E. Singer, Deanna J. Reichel, Geoff D. Biegler, Limin Zheng, Elizabeth M. Flanagan

Parsons Behle & Latimer: David G. Mangum, C. Kevin Speirs, Kristine Edde Johnson, Michael R. McCarthy

Myriad Genetics, Inc.: Benjamin G. Jackson, Matthew Gordon

Date: April 18, 2014

/s/Jonathan E. Singer

Jonathan E. Singer

UNIVERSITY OF UTAH, et al. v. AMBRY GENETICS

14-1361, -1366

CERTIFICATE OF INTEREST

Counsel for the Plaintiffs-Appellants, University of Utah Research Foundation, The Trustees of the University of Pennsylvania, HSC Research and Development Limited Partnership, and Endorecherche, Inc. certifies the following:

1. The full name of every part or amicus represented by me is:

University of Utah Research Foundation, The Trustees of the University of Pennsylvania, HSC Research and Development Limited Partnership, and Endorecherche, Inc.

2. The name of the real party in interest represented by me is:

University of Utah Research Foundation, The Trustees of the University of Pennsylvania, HSC Research and Development Limited Partnership, and Endorecherche, Inc.

3. All parent corporations and any publicly held companies that own 10 percent or more of the stock of the party or amicus curiae represented by me are:

None.

4. The names of all law firms and the partners or associates that appeared for the party or amicus now represented by me in the trial court or agency or are expected to appear in this court are:

Parsons Behle & Latimer: David G. Mangum, C. Kevin Speirs, Kristine Edde Johnson, Michael R. McCarthy

Date: April 18, 2014

/s/David G. Mangum

David G. Mangum

TABLE OF CONTENTS

	<u>Page</u>
CERTIFICATE OF INTEREST	
TABLE OF CONTENTS	i
TABLE OF AUTHORITIES	iii
STATEMENT OF RELATED CASES.....	1
JURISDICTIONAL STATEMENT	1
STATEMENT OF THE ISSUES	1
STATEMENT OF THE CASE	3
STATEMENT OF FACTS	6
I. TECHNOLOGY BACKGROUND	6
A. Genes Are Functional Units of DNA That Store and Transmit Biological Information	6
B. Scientists Use Man-Made Tools that Take Advantage of Watson-Crick Base Pairing to Study Genes, Including PCR and Single-Stranded DNA Primers, DNA Probes and cDNAs.	8
II. THE DISCOVERY OF THE BRCA GENES AND MYRIAD’S BRACANALYSIS® TEST.....	12
III. THE PATENTS ON APPEAL AND PRECEDING AMP LITIGATION	16
IV. THE LITIGATION AGAINST AMBRY	19
SUMMARY OF THE ARGUMENT	25
ARGUMENT	27
I. LEGAL STANDARDS.....	27

TABLE OF CONTENTS (cont'd)

	<u>Page</u>
II. THE DISTRICT COURT ERRED IN FINDING A SUBSTANTIAL QUESTION OF PATENT ELIGIBILITY OF THE METHOD CLAIMS	28
A. Representative Claims 7 and 8 Are Patent-Eligible Methods of Applying the BRCA1 DNA Sequence.....	28
B. Representative Claims 7 and 8, Properly Viewed as a Whole, Contain an “Inventive Concept” Under Mayo.....	32
C. Representative Claims 7 and 8 Do Not Unlawfully Preempt the BRCA1 DNA Sequence	37
III. THE DISTRICT COURT ERRED IN FINDING A SUBSTANTIAL QUESTION OF PATENT ELIGIBILITY OF THE PRIMER PAIR CLAIMS	45
A. A Composition of Matter Is Patent Eligible under Section 101 and Supreme Court Precedent Unless it is a “Product of Nature”—i.e., Both its Structure and Utility Are as in Nature, Unaltered by Man.....	45
B. The Primer Pair Claims Are Not a Product of Nature	48
1. The Primer Pair Claims Are Designed and Made by Scientists, Not Snipped from Nature	48
2. The Claimed Primer Pairs Have Utility Distinct from Natural DNA or Fragments Thereof.....	50
C. The District Court Vastly Overread AMP to Include Primers and Probes as Patent Ineligible.....	55
IV. THE DISTRICT COURT ABUSED ITS DISCRETION IN ANALYZING THE BALANCE OF HARSHIPS AND PUBLIC INTEREST FACTORS	59
CONCLUSION.....	61
CERTIFICATE OF COMPLIANCE	64

TABLE OF AUTHORITIES

Page(s)

CASES

<i>Abbott Labs. v. Sandoz, Inc.</i> , 544 F.3d 1341 (Fed. Cir. 2008)	27
<i>Ariosa Diagnostic, Inc. v. Sequenom, Inc.</i> , No. C-11-06391, 2013 WL 5863022 (N.D. Cal. Oct. 30, 2013)	23, 43, 44
<i>Ass’n for Molecular Pathology v. Myriad Genetics, Inc.</i> , 133 S. Ct. 2107 (2013)	<i>passim</i>
<i>Ass’n for Molecular Pathology v. Myriad Genetics, Inc.</i> , 689 F.3d 1303 (Fed. Cir. 2012)	<i>passim</i>
<i>Bilski v. Kappos</i> , 130 S. Ct. 3218 (2010)	28, 29
<i>In re BRCA1- and BRCA2- Based Hereditary Cancer Test Patent Litigation</i> , MDL Case No. 2:14-MD-2510	1
<i>Celsis In Vitro, Inc. v. Cellzdirect Inc.</i> , 664 F.3d 922 (Fed. Cir. 2012)	61
<i>CLS Bank Int’l v. Alice Corp.</i> , 717 F.3d 1269 (Fed. Cir. 2013)	28, 42
<i>Diamond v. Chakrabarty</i> , 447 U.S. 303 (1980)	46, 48, 54
<i>Diamond v. Diehr</i> , 450 U.S. 175 (1981)	<i>passim</i>
<i>Funk Bros. Seed Co. v. Kalo Inoculant Co.</i> , 333 U.S. 127 (1948)	24, 45, 54

TABLE OF AUTHORITIES (cont'd)

	<u>Page(s)</u>
<i>Gottschalk v. Benson</i> , 409 U.S. 63 (1972)	29, 40, 41
<i>Harris Corp. v. Ericsson Inc.</i> , 417 F.3d 1241 (Fed. Cir. 2005)	27
<i>Koon v. United States</i> , 518 U.S. 81 (1996)	27, 59
<i>Mayo Collaborative Services v. Prometheus Laboratories, Inc.</i> , 132 S. Ct. 1289 (2012)	<i>passim</i>
<i>O'Reilly v. Morse</i> , 56 U.S. 62 (1854)	40, 41
<i>Parker v. Flook</i> , 437 U.S. 584 (1978)	22, 33, 42
<i>Pfizer, Inc. v. Teva Pharms., USA, Inc.</i> , 429 F.3d 1364 (Fed. Cir. 2005)	61
<i>Robert Bosch LLC v. Pylon Mfg. Corp.</i> , 659 F.3d 1142 (Fed. Cir. 2011)	27
<i>Sanofi-Synthelabo v. Apotex, Inc.</i> , 470 F.3d 1368 (Fed. Cir. 2006)	60
<i>The Telephone Cases</i> , 8 S. Ct. 778 (1888)	41
<i>Titanium Metals Corp. v. Banner</i> , 778 F.2d 775 (Fed. Cir. 1985)	57
<i>Trebo Mfg., Inc. v. Firefly Equip., L.L.C.</i> , -- F.3d --	27

TABLE OF AUTHORITIES (cont'd)

	<u>Page(s)</u>	
<i>Ultramercial, Inc. v. Hulu, LLC</i> , 722 F.3d 1335 (Fed. Cir. 2013)	28, 37, 38	
 STATUTES		
28 U.S.C. § 1292(a)(1) and (c)(1)	1	
28 U.S.C. §§ 1331, 1338 and 2201	1	
35 U.S.C. § 101	<i>passim</i>	
 OTHER AUTHORITIES		
 GUIDANCE FOR DETERMINING SUBJECT MATTER ELIGIBILITY OF CLAIMS RECITING OR INVOLVING LAWS OF NATURE, NATURAL PHENOMENA & NATURAL PRODUCTS 11, <i>available at</i> http://www.uspto.gov/patents/law/exam/myriad- mayo_guidance.pdf		44
Ronald Bailey, <i>Warning: Bioethics May Be Hazardous to Your Health</i> , Aug. 1, 1999, http://reason.com/archives/1999/08/01/warning- bioethics-may-be-hazar/print	15	

STATEMENT OF RELATED CASES

In addition to this case, Appellants (collectively, “Myriad”) have filed lawsuits concerning the same patents and claims against a number of other defendants that will likely be affected by the result in this appeal. All those lawsuits have been consolidated along with this case into an MDL proceeding before the same district court, captioned *In re BRCA1- and BRCA2- Based Hereditary Cancer Test Patent Litigation*, MDL Case No. 2:14-MD-2510.

JURISDICTIONAL STATEMENT

The district court has jurisdiction over this case under 28 U.S.C. §§ 1331, 1338 and 2201. On March 10, 2014, the district court issued its Memorandum Decision and Order Denying Myriad’s Motion for Preliminary Injunction. Myriad filed its notice of appeal on March 13, 2014. This Court has jurisdiction over this appeal under 28 U.S.C. § 1292(a)(1) and (c)(1).

STATEMENT OF THE ISSUES

1. Whether the district court erred in finding that Ambry raised a substantial question that Myriad’s method claims are not directed to patent eligible subject matter under 35 U.S.C. § 101, where those claims employ previously unperformed steps involving a newly-discovered gene sequence and tools designed specifically to utilize that sequence; and

2. Whether the district court erred in finding that Ambry raised a substantial question that Myriad's pair of primer claims are not directed to patent eligible subject matter under 35 U.S.C. § 101, where those claims are directed to a pair of single-stranded DNA primers for use in PCR that do not occur in nature and have utility engineered by man that is different from naturally-occurring DNA.

3. Whether the district court erred in finding that the balance of hardships favored the denial of an injunction, where its analysis relied on its erroneous finding of a substantial question of invalidity.

4. Whether the district court erred in finding that the public interest did not favor the entry or denial of an injunction, despite this Court's longstanding acknowledgment of the public's strong interest in upholding patent rights to promote research and development aimed at treating disease.

5. Whether the district court's denial of a preliminary injunction, in light of its clearly erroneous findings on patent eligible subject matter, constituted an abuse of discretion where it found that Myriad was likely to be irreparably harmed in the absence of an injunction.

STATEMENT OF THE CASE

In the early 1990s, scientists from Myriad and its academic research partners discovered the precise location and sequence of BRCA1, the first gene linked to hereditary breast and ovarian cancer. This landmark discovery was hailed worldwide, even leading the NBC Nightly News on the day it was announced. “An extraordinary advance for cancer research,” the *Wall Street Journal* called it, “opening the way for new and powerful ways to diagnose and treat breast cancer early.” True to the *Journal*’s prediction, the discovery led to a first of its kind diagnostic test, Myriad’s BRACAnalysis®, which for the first time gave to patients the ability to assess accurately their risk for these too-often deadly diseases. Since its launch in 1996, more than one million patients have benefited from BRACAnalysis®, one of the most extensively utilized genetic tests in the United States.

But the widespread availability of this test did not come without an extraordinary price—Myriad had to invest in excess of 500 million dollars not only to discover the BRCA genes, but to develop reliable diagnostic methods and tools to apply that discovery, and to educate physicians, patients, health care providers, and insurers on the utility of a diagnostic test for a “common” disease. Over great skepticism, Myriad pioneered the market for genetic breast/ovarian cancer testing, so much so that BRACAnalysis® has now become the standard of care.

None of this would have been possible without the asserted claims that are the subject of this appeal, and the exclusivity they once afforded. These claims cover the diagnostic methods and tools—BRCA-related probes and primers—that Myriad employs in conducting BRCAAnalysis®, and are plainly infringed by Ambry and other new entrants to the diagnostic breast cancer market. But despite finding that Myriad would be irreparably harmed by Ambry’s actions, the district court refused to enjoin Ambry during the pendency of Myriad’s infringement suit because, according to the district court, not a single one of Myriad’s asserted patent claims was likely valid under 35 U.S.C. § 101.

This appeal thus presents this Court with one of the most basic issues under patent law: should Myriad be able to enforce its patents to enjoy its remaining exclusivity or does section 101 of the Patent Act effectively say that Myriad can patent almost nothing related to its ground-breaking discovery?

This Court should resolve these questions resoundingly in Myriad’s favor. While the Supreme Court may have ruled that the isolated BRCA genes themselves were not patentable, it declined to go any further, and stated that Myriad could lawfully claim applications of that gene, including, for example, BRCA cDNAs. It is thus manifestly not the case—as the district court found—that Myriad’s methods of diagnosis employing BRCA-related probes and primers are “conventional” or unlawfully preempt use of the gene, and it is manifestly not the

case that pairs of short, single-stranded BRCA-related DNA primers created in the lab to initiate the polymerase chain reaction (“PCR”) in order to replicate portions of the BRCA genes are unpatentable “products of nature” under the Supreme Court’s precedent.

The district court’s opinion gives breathtaking scope to the judicially-created “exceptions” to section 101, expanding them to the point where they threaten to swallow not only section 101, but large portions of sections 102 and 103 as well. According to the district court, because Myriad did not invent, separate from BRCA-related tools or methods employing such tools, a new method of analyzing genes or new tools to analyze genes, Myriad cannot patent any claim having an “inventive concept” that is tied in some fashion to the BRCA genes.

This is not what this Court’s and the Supreme Court’s authorities provide. Rather, the Supreme Court’s decision in *Association for Molecular Pathology v. Myriad Genetics, Inc.*, 133 S. Ct. 2107, 2120 (2013) (“AMP”), provides that, as the discoverer of the BRCA genes, Myriad “was in an excellent position to claim applications of [its] knowledge” of the BRCA sequences, citing with approval a Judge Bryson opinion from this Court identifying a claim in the asserted ’441 patent that is very similar to the method claims at issue here as an example of a patentable application of that knowledge. That is what Myriad’s asserted claims do, and what they cover—applications of the knowledge of the BRCA genes

sequence. They plainly meet the requirements of both *AMP* and *Mayo Collaborative Services v. Prometheus Laboratories, Inc.*, 132 S. Ct. 1289 (2012), and are patentable under section 101. The district court's contrary determination should be reversed.

STATEMENT OF FACTS

I. TECHNOLOGY BACKGROUND

Because this Court is generally familiar with genes and DNA from previous cases, *e.g.*, *Association for Molecular Pathology v. Myriad Genetics, Inc.*, 689 F.3d 1303 (Fed. Cir. 2012), Myriad provides an abbreviated overview of the relevant technology, with additional detail provided on the biological processes and molecular biology tools implicated by the asserted claims.

A. Genes Are Functional Units of DNA That Store and Transmit Biological Information

The terms “native DNA,” “genomic DNA” or “gDNA” refer to DNA as it exists in nature in the human body. [A6797; A6801-02.] *AMP*, 689 F.3d at 1313. In brief, native DNA is composed of hundreds of millions of units of nucleotides, commonly referred to by the nucleotide bases adenine (A), thymine (T), guanine (G) and cytosine (C). *AMP*, 689 F.3d at 1310. DNA generally takes the form of a double-stranded helix, which consists of two intertwined strands of DNA in which the bases of the nucleotides from one strand are bonded to the bases of the nucleotides from the other strand in a specific, complementary fashion known as

Watson-Crick base pairing. *Id.* at 1310-11. [*See also* A6800-01; A6298; A6304-05.]

Every gene has a unique DNA sequence—represented by the linear order of its nucleotide bases—that includes protein coding regions (exons) and non-coding regions (introns). *AMP*, 689 F.3d at 1311. [A6801-02; A6304; A6311-15; A1952; A1958-59; A5490; A5506.] To generate a protein from a gene, DNA is first transcribed and processed into messenger RNA (mRNA), which contains only sequences complementary to the exons of the DNA, and then the mRNA is translated into a protein. *AMP*, 689 F.3d at 1311-12. [A6803-05; A6313-15.]

The human body synthesizes, or “replicates,” DNA every time a new cell is generated. [A6310.] In that process, the DNA double helix unwinds, and each strand then serves as a template to synthesize a new, complementary strand that is a replica of the opposite strand. [A6309-11.] Thus, from one DNA double helix is generated two DNA double helices. This DNA replication process is initiated or “primed” by short strands of RNA, known as RNA “primers,” that are complementary to portions of the template strands of DNA and serve as the starting material for DNA synthesis. [A7611; A7616-17.] DNA replication is catalyzed by DNA polymerase, an enzyme that adds nucleotides, one at a time, to the free 3’ end of the primer and the growing strand, elongating it in an order

complementary to the template DNA to generate a new strand of DNA.¹ [A6309-11; A7616-17; A7623.] DNA polymerase can only add nucleotides in a 5' to 3' direction because it can only add the 5' phosphate of a new nucleotide to a free 3'-hydroxyl (“-OH”) group of an existing nucleotide sequence. [A6309-11; A7616-17.] There are no short, single strands of DNA with a free 3'-OH group in nature that can serve as primers. In natural DNA replication, RNA primers are used as the starting material. [A7616-17.]

For each gene there typically exists a sequence and certain variations that most commonly occur, typically called the “wild-type” sequence. [A1964.] Some variants from the wild-type sequence are harmless, but others—mutations—can cause disease. [A6802.] A “germline” sequence is the sequence that a person has at birth, which may or may not vary from the “wild-type.” [A1964; A5506.]

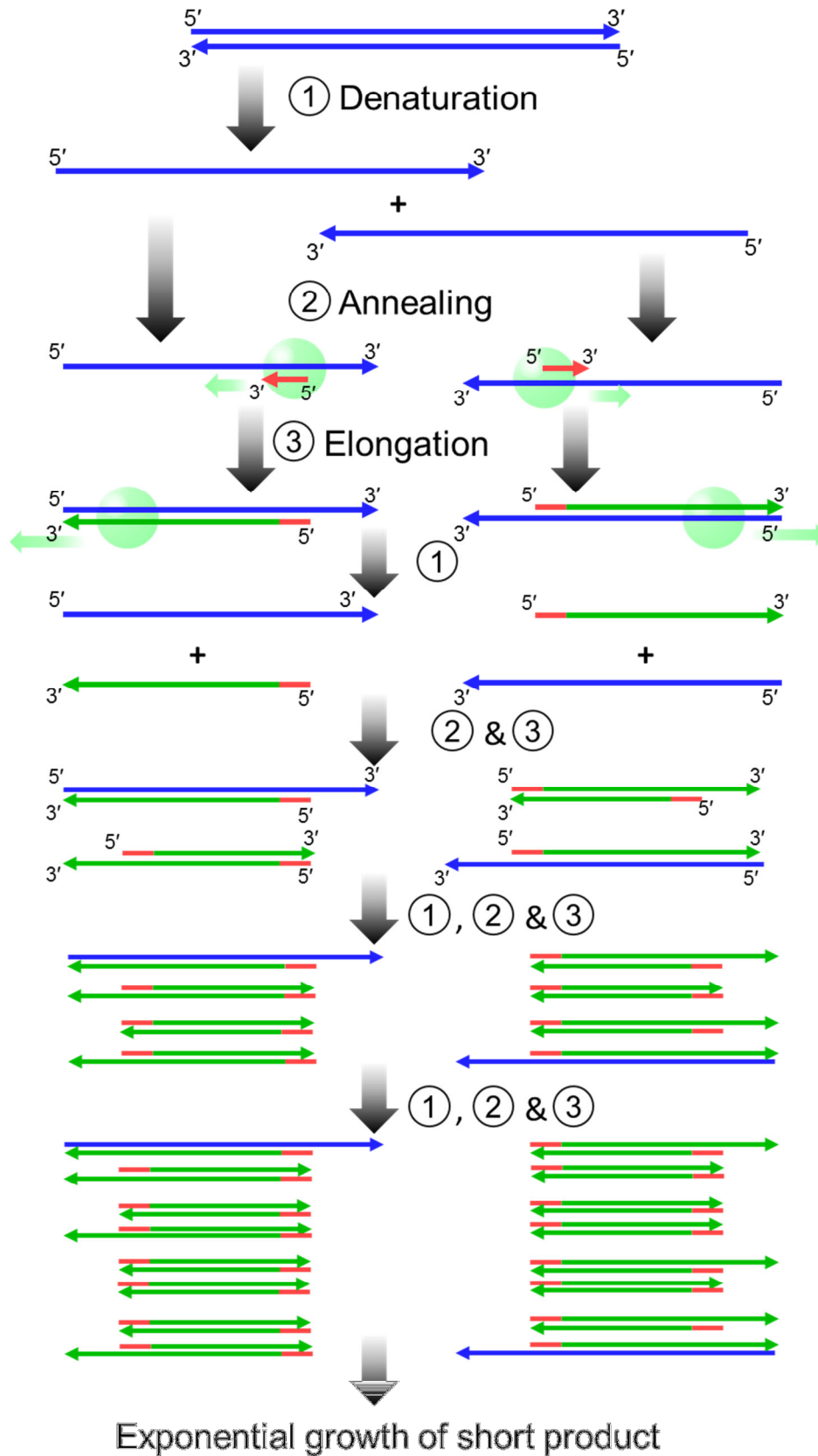
B. Scientists Use Man-Made Tools that Take Advantage of Watson-Crick Base Pairing to Study Genes, Including PCR and Single-Stranded DNA Primers, DNA Probes and cDNAs.

Over the years, scientists have developed techniques to synthesize and utilize DNA in the lab to study genetic diseases and the mutations that can cause them. [A6806.] Most relevant to this appeal is PCR, the general process of which was invented in the mid-1980s. PCR typically uses an input DNA sequence (such

¹ The two ends of a single-stranded DNA are considered to be chemically different. [A6802.] The “beginning” of a DNA molecule is the 5' end, and the “end” of a DNA molecule is the 3' end. [*Id.*]

as native DNA isolated from the human body), a thermostable DNA polymerase enzyme, a pool of all four DNA nucleotides, and a large excess of two distinct single-stranded DNA primers. [A6807.] These primers are designed to be complementary to discrete regions of the input DNA flanking the portion of the DNA in interest, in order to make copies of that portion of the input DNA. [*Id.*]

PCR involves several steps that are cyclically repeated to generate synthetic copies of the input DNA in interest. [A6807.] First, the reaction mixture is heated, which causes the DNA double-helix to separate into single strands that serve as templates. [*Id.*] Then the reaction mixture is cooled, which permits each primer to bind to the complementary portion of its single-stranded DNA template. [*Id.*] Next, the DNA polymerase enzyme adds nucleotides to the free 3' end of the elongating strand, starting with the primer, in an order complementary to the template DNA, generating a complementary copy of the template DNA. [*Id.*] These steps are then repeated to generate numerous synthetic copies of the target input DNA sequence between the two primers. In this fashion, PCR results in an amplified PCR product that is double-stranded and incorporates the sequence of the first primer followed by the target sequence, and further followed by the sequence of the second primer, as shown below. [A6808; *see also* A1955-56.]



[A1955-56 (blue represents isolated DNA fragments, green represents synthetic DNA molecule, and red represents synthetic DNA primers).]

In a regular PCR, the primers are typically pairs of artificial single-stranded DNA molecules, 15 to 30 nucleotides in length. [A1954-62; A6806; A6332.] Primer pairs are specific to the target DNA region they are designed to amplify. Each primer must uniquely hybridize to a segment of DNA bordering the target DNA region; collectively, the primer pair defines the borders of the target DNA. A primer pair designed to be specific for one region of a gene will not be effective in priming PCR for a different gene or a different target region within the same gene. [A1954-55; A1960-62; A6318-26; A6806-07.]

Probes are another man-made tool that can be used in the lab to investigate an input DNA sequence. Like primers, they are designed and synthesized to be complementary to, and thus hybridize to, a discrete portion of a target DNA. Probes are generally longer than primers and include a label, and are typically used to identify a DNA segment of interest and not to create copies of it. [A6331.]

Scientists can also make a type of DNA called complementary DNA or cDNA. *AMP*, 689 F.3d at 1313. cDNA is complementary to the mRNA from which it is synthesized, and thus includes the exons but lacks the introns in native DNA. *Id.* at 1313-14. [A6805-06; A6329.]

II. THE DISCOVERY OF THE BRCA GENES AND MYRIAD'S BRACANALYSIS® TEST

The story of the discovery of the BRCA genes dates back to the 1960s when Dr. Mark Skolnick, a co-founder of Myriad, first took an interest in genetics. [A6911-12.] At that time, Dr. Skolnick had the unique insight to apply the study of human populations—a field typically considered in the context of sociology and economics—to genetics. [A6912.] Joining with a prominent geneticist in the field, Dr. Skolnick learned how to construct genealogies, subsequently creating the first computerized genealogical record. [A6912-13.] Shortly thereafter, Dr. Skolnick learned of the Utah Genealogical Society and, with his team, radically reconstructed the entire Utah Mormon Genealogy in order to link it to the Utah Cancer Registry to study familial patterns of cancer—an achievement recognized by the NIH. [A6913.] In tandem with these efforts, Dr. Skolnick and his team also set up a familial cancer screening clinic, studying tens of thousands of family members with a history of cancer, including breast and ovarian cancers. [A6914.]

Dr. Skolnick was also a pioneer in molecular biology. Among other things, his group developed an important technique for genetic mapping—Restriction Fragment Length Polymorphisms. [A6914.] Using that technique and others, his team successfully mapped the gene for neurofibromatosis and mapped and cloned the gene for Alport Syndrome. [A6914.]

Thus, when it was announced in late 1990 that a gene, coined as BRCA1, linked to chromosome 17q21 was potentially responsible for a large number of familial breast and ovarian cancers [A6809; A6915], Dr. Skolnick set out to find it. He organized a large team of academic, government, and industry scientists from across the United States and Canada, and, because of the insufficient grants available for the project, funded it by creating Myriad in 1991 in conjunction with a venture capitalist. [A6915.]

Dr. Skolnick's familial database was instrumental in this search. Because the BRCA1 protein had not yet been identified, a positional cloning approach—which relies on linkage analyses of familial data to search for the location of the disease-causing gene—was one of the techniques used to find it.² [A6809; A6879-81.] Along with other molecular biology techniques, Dr. Skolnick's team used his database to home in on the BRCA1 gene—which is now known to extend over only about 100,000 base pairs—starting from a region of millions of base pairs. [A6809-11; A6911-16; A6882.] This was no easy task—other groups of researchers undertook similar efforts to find BRCA1, pointing to various, and oftentimes divergent, areas on chromosome 17q21 within which they hypothesized the BRCA1 gene might be located. [A6881-90.]

² If the BRCA1 protein sequence were known, researchers might have used other techniques to work backwards from the protein to the mRNA, and ultimately the gene. [A6809.]

Ultimately, Dr. Skolnick's team discovered the location and sequence of the BRCA1 gene, and later BRCA2. *AMP*, 133 S. Ct. at 2112. This discovery was deservedly hailed world-wide, described on the day it was announced by the NBC Nightly News as "one of the most important breakthroughs in breast cancer research in years." [A7234-35; A27.] It was truly a "landmark" discovery and "scientific triumph," and was published in the most respected journals in the scientific community. [See A7238-44; A7376-86.]

But the discovery of the genes, in and of themselves, yielded nothing—without a reliable diagnostic test applying that knowledge, and people willing to invest in it, there is no telling what might have resulted. [A7398 ("[T]he public should know applying [the gene] will involve as much imagination and hard work as it took to find it.").] While such a test may be thought of as routine today, in the 1990s it was not—Myriad's development of the test for the BRCA genes was the first such test for a so-called "common" disease. [A1936-37.] Fortunately, Myriad was willing to undertake this risky endeavor, supported by the knowledge that the tools used to run the test—DNA primers and probes and methods of using the same—were protected from immediate copying by the patents on appeal here. Myriad spent tens upon tens of millions of dollars to develop the techniques and run clinical trials to apply its discovery, and then hundreds of millions more to educate the medical community about it. [A1938; A7259-60.] Skeptics abounded,

including those who asserted that women should not be provided with the information revealed by the test, as the secondary sources describe. See Ronald Bailey, *Warning: Bioethics May Be Hazardous to Your Health*, Aug. 1, 1999, <http://reason.com/archives/1999/08/01/warning-bioethics-may-be-hazar/print>.

The result of all of this effort is Myriad's BRACAnalysis® test, the first version of which was launched in 1996. [A1938.] As currently configured, this test detects the presence of and characterizes mutations in the BRCA genes that are responsible for the majority of hereditary breast and ovarian cancers. [A1937.] Since its launch, over one million patients have benefited from BRACAnalysis®, which is one of the most widely employed genetic tests for disease in the United States. [A1937.]

Myriad's integrated test has a current list price of about \$4,000, which today is typically covered by insurance because of Myriad's efforts to make that happen. [A1938-39.] For patients who cannot afford the test, Myriad has set up numerous patient assistance programs to help pay for it. [A10354-55.] And while Myriad's opponents constantly tout the total amount of revenue that Myriad has generated from the test—roughly \$2 billion in total at the time of the preliminary injunction proceedings—the company itself lost money every year from its founding in 1991 until 2007.

III. THE PATENTS ON APPEAL AND PRECEDING *AMP* LITIGATION

As the Court is well aware, Myriad’s patents related to its discovery of BRCA genes have been the subject of extensive litigation. In brief, in 2009, the Association for Molecular Pathology (“AMP”), supported by the ACLU, sought a declaratory judgment that certain of these patents’ claims were invalid under section 101. The claims challenged in that litigation were directed to (1) “isolated” DNA, which included DNA isolated from the natural BRCA genes, (2) isolated BRCA cDNAs only, and (3) general methods for “analyzing” or “comparing” a patient’s BRCA sequence with the natural sequence for the genes. *AMP*, 689 F.3d at 1309-10. Neither the composition claims directed to primers or probes, nor the method claims that include any physical steps—*i.e.*, all those at issue now—were challenged.

After Judge Sweet invalidated all the claims raised by AMP, on appeal and after an initial remand from the Supreme Court in light of the *Mayo* case, this Court found that both the isolated DNA and cDNA composition claims were patent eligible under section 101, but that the “analyzing” and “comparing” method claims at issue there were not eligible because they recited no physical steps and claimed only “abstract mental processes.” *Id.* at 1333-35.

In an opinion concurring-in-part and dissenting-in-part in that judgment, Judge Bryson noted that, “[o]f course, Myriad is free to patent applications of its

discovery. As the first party with knowledge of the sequences, Myriad was in an excellent position to claim applications of that knowledge. Many of its unchallenged claims are limited to such applications.” *Id.* at 1349. He further cited several of the “unchallenged claims” as examples of patent-eligible applications of knowledge, including one claim—claim 21 of the ’441 patent—that is strikingly similar to the exemplary method claims at issue in this appeal, as shown in the table below. *Id.*

'441 patent claims 7 and 8	'441 patent claim 21
7. The method of claim 1 wherein a germline nucleic acid sequence is compared by hybridizing a BRCA1 gene probe which specifically hybridizes to a BRCA1 allele to genomic DNA isolated from said sample and detecting the presence of a hybridization product where a presence of said product indicates the presence of said allele in the subject.	21. The method of claim 20 wherein a germline alteration is detected by hybridizing a BRCA1 gene probe which specifically hybridizes to an allele of one of said alterations to RNA isolated from said human sample and detecting the presence of a hybridization product, wherein the presence of said product indicates the presence of said allele in the sample.
8. The method of claim 1 wherein a germline nucleic acid sequence is compared by amplifying all or part of a BRCA1 gene from said sample using a set of primers to produce amplified nucleic acids and sequencing the amplified nucleic acids.	

On appeal to the Supreme Court, only the isolated DNA and cDNA composition claims were at issue; Myriad did not appeal the general method claims that had been invalidated as abstract ideas. But even as to the composition claims,

petitioners emphasized at oral argument that the case should not impact claims to primers and probes. *AMP* Oral Argument Tr. (Apr. 15, 2013) at 9-10, available at http://www.supremecourt.gov/oral_arguments/argument_transcripts/12-398_h3dj.pdf. When Justice Sotomayor asked whether “[t]he primers and probes stand,” counsel for the petitioners agreed that the primer and probe claims “would still remain. Even if you were to rule for Petitioners, you would not have to rule concerning the use of DNA as a probe or a primer.” *Id.* at 9-11.

On June 13, 2013, the Supreme Court issued its decision, finding that the composition claims directed to isolated DNA (or fragments thereof) would cover “products of nature” and were not patentable just “because they have been isolated from the surrounding genetic material.” *AMP*, 133 S. Ct. at 2120. Because those claims read in part on unaltered, genomic DNA removed from the body, they were unpatentable products of nature. *Id.* at 2117-18.

The Court explained, though, that “cDNA does not present the same obstacles to patentability” because it is not “naturally occurring,” even though nature might create a cDNA in a “pseudogene.” *Id.* at 2119 n.8. This was so, even though the sequence of cDNA is, in fact, “dictated by nature.” *Id.* at 2119.

After making these twin holdings, the Court took the time to note what was not at issue in the case, writing that the “case d[id] not involve patents on new applications of knowledge about the BRCA1 and BRCA2 genes.” *Id.* at 2120. Of

Myriad's ability to claim such applications, the Court stated as follows: "Judge Bryson aptly noted that: '[a]s the first party with knowledge of the sequences, Myriad was in an excellent position to claim applications of that knowledge. Many of its unchallenged claims are limited to such applications.'" *Id.* (quoting *AMP*, 689 F.3d at 1389).

IV. THE LITIGATION AGAINST AMBRY

The same day the Supreme Court ruled in *AMP*, Ambry announced that it would begin offering its own BRCA testing to compete with Myriad's. Taking advantage of the market that Myriad had built, Ambry offered its tests at a price severely undercutting Myriad's. After analyzing Ambry's proposed product offerings and concluding there was infringement, Myriad sued Ambry in the United States District Court for the District of Utah on July 9, 2013.

That same day, Myriad moved for a preliminary injunction to prevent the irreparable harm it would surely suffer if Ambry remained on the market, including price erosion, loss of market share and its investments in developing the market, and loss of its patent exclusivity. Myriad's motion raised ten claims from six patents. For purposes of this appeal, four of these claims are representative: claims 7 and 8 of the '441 patent and claims 16 and 17 of the '282 patent.³

³ Asserted primer claims 29 and 30 of the '492 patent are similar to claims 16 and 17 of the '282 patent, the primary difference being that they concern BRCA2

Representative claims 7 and 8 of the '441 patent claim methods for screening germline for mutations (“alterations”) in the BRCA1 gene. They depend from claim 1, one of the claims invalidated by this Court as claiming merely an abstract idea. But unlike claim 1, they require specific steps to make the comparison. In claim 7, these steps are hybridizing a BRCA1 gene probe to a person’s native BRCA1 allele and detecting the presence of the hybridization product to indicate the presence of the allele. In claim 8, the steps are amplifying the BRCA1 gene through the use of a set of primers and sequencing the amplified DNA. [A250-351 at A348.]

Claims 16 and 17 of the '282 patent cover pairs of single-stranded DNA primers for use in PCR. Claim 16 is directed to primer pairs used to amplify a portion of the sequence of the entire BRCA1 gene, while claim 17 is limited to primer pairs used to amplify a portion of the sequence of the BRCA1 cDNA (*i.e.*, exons only). [A146-249 at A242.]

Ambry opposed the motion for preliminary injunction in an avalanche of paper, arguing that its products do not infringe and that all the asserted claims are invalid under 35 U.S.C. §§ 101, 102, 103 and 112. Ambry also denied that Myriad

primers instead of BRCA1. [A242; A446.] Asserted method claims 2 and 4 of the '155 patent are similar to claims 7 and 8 of the '441 patent in that they employ primers to achieve their methods. [A348; A140-41.] Appellants are no longer relying on asserted claim 5 of the '721 patent and asserted claim 4 of the '857 patent for purposes of seeking preliminary injunctive relief.

would be irreparably harmed in the absence of an injunction, and argued that the balance of hardships and the public interest favored denial of an injunction.

The district court heard oral argument and received limited testimony on the motion over portions of three days in the fall of 2013. On March 10, 2014, the district court denied Myriad's motion. As to irreparable harm, the district court found that Myriad had presented "logical and persuasive testimony" that it "will suffer irreparable financial harm if an injunction does not issue," and that damages would not be adequate to compensate for this harm. [A61-68.]

However, the district court also found that Ambry had raised a substantial question that the asserted claims were invalid under 35 U.S.C § 101. On the method claims, the district court did not address them separately but, as a group, found that they bore a "striking" similarity to the abstract method claims invalidated in *AMP*, without any discussion of claims' 7 and 8 near identity to the claims identified by Judge Bryson as patentable applications of the BRCA knowledge. [A90.]

After making these comments, the district court then summarized the analysis it believed was required by *Mayo* as follows: "whether the Method Claims at issue set forth an 'inventive step' aside from the patent ineligible subject matter, and beyond 'well-understood, routine, conventional activity previously engaged in' by those in the field; and 2) whether allowing the Method Claims risks

preempting the use of a natural law, natural phenomenon, or abstract idea.” [A93.]

In describing the question in this fashion, the district court heavily emphasized *Parker v. Flook*, 437 U.S. 584 (1978), which treats unpatentable subject matter separately as part of the prior art, without acknowledging that that portion of *Flook* was overruled by *Diamond v. Diehr*, 450 U.S. 175 (1981). [A92.]

The district court then went on to conduct the analysis it asserted was required. In so doing, however, it divided the method claims up into separate pieces, contrary to the *Diehr* case it had just discussed. Ignoring the BRCA gene sequences entirely, the district court found that “the other steps set forth in the Method Claims are conventional activities that were well-understood and uniformly employed by those working with DNA at the time Myriad applied for its patents.” [A94.] In other words, according to the district court, because Myriad had not invented some new way of assessing mutations or genes on top of its discovery of the BRCA sequences, it could not patent particular applications of the BRCA sequences, such as new methods of testing a patient’s sample, using new primers and probes, to determine the patient’s BRCA sequence. Because “[t]he laboratory materials, reagents, and protocols” to accomplish “amplification, sequencing, comparisons, detecting alterations in sequences, and hybridizing probes to alleles” were “well known and widely available in the art,” in general, the district court held that Myriad’s specific application of those techniques to the

unknown BRCA1 sequences did not satisfy section 101. [A94.]

On preemption, the district court found that, “if allowed, Plaintiffs’ Method Claims would essentially foreclose the most widely used means to study and test for BRCA1 and BRCA2 genes” because “PCR using primers and probe hybridization are the means needed to determine and compare BRCA1 and BRCA2 sequences, and to conduct BRCA1 and BRCA2 tests.” [A95-96.] The district court cited no decision of this Court or the Supreme Court for its “most widely used means” test, but instead relied almost exclusively on the analysis from another district court—*Ariosa Diagnostic, Inc. v. Sequenom, Inc.*, No. C-11-06391, 2013 WL 5863022 (N.D. Cal. Oct. 30, 2013). [A96-100.] In reaching its preemption conclusion, the district court did not connect its analysis with whether or not the use of the unpatentable subject matter—the native BRCA gene sequences—inherently requires the claimed steps of the asserted method claims, as the Supreme Court did in *Mayo*.

On the primer claims, the court spent most of its time addressing the issue of whether or not the Supreme Court’s *AMP* ruling affirmed patentability of any DNA that was synthetic, an argument Myriad had raised, and gave only limited consideration to the specific claims at issue that are limited to primer pairs for use in PCR. In answering the first question in the negative—that synthetic DNA may be unpatentable under the Supreme Court’s ruling—the district court read the

Supreme Court’s *AMP* holding very broadly, interpreting it to reach any claim that “reflects” naturally-occurring sequences despite the Supreme Court’s statement that cDNA was patentable even though its sequence was “dictated” by nature. [A75-81.] Remarkably, at the behest of Ambry, the district court concluded that primers and probes were, in fact, at issue in the prior litigation despite the opposite clearly being the case. [*Id.*]

In its limited analysis of the actual substance of the primer pair claims, the district court focused most of its attention on whether those claims are “directed to compositions structurally similar to the DNA found in nature.” [A83.] The court spent only one paragraph discussing Myriad’s argument that the claimed primers are patent-eligible because they are functionally different than natural DNA, concluding that the primers’ utility—their use in PCR—“exploits [a] natural DNA function to a useful end” and is not “markedly different” from natural DNA. [A87.] The court similarly dismissed Myriad’s argument that primer pairs inarguably do not exist in nature as an argument “reminiscent” of that rejected in *Funk Bros. Seed Co. v. Kalo Inoculant Co.*, 333 U.S. 127 (1948), because “each primer [is] identical to the BRCA1 or BRCA2 nucleotide sequence from which it is derived, and carr[ies] identical genetic information.” [A85.] That PCR does not occur in the human body did not matter to the district court because, “during PCR, the primers function similarly to genomic DNA undergoing replication in the

human body.” [*Id.*]

Based on all of the above, the district court found that Myriad was unlikely to succeed on the merits of its claim because each of its asserted claims was likely invalid under section 101. Accordingly, the court denied Myriad’s motion for preliminary injunction.⁴ This appeal followed.

SUMMARY OF THE ARGUMENT

The district court abused its discretion in denying Myriad’s motion for a preliminary injunction. On the method claims, the district court grouped them together and failed to analyze them in a meaningful fashion. In so doing, the court erroneously found that they are “strikingly” similar to the method claims invalidated as too abstract in *AMP*, ignoring their specific physical limitations that readily distinguish them from those claims, as Judge Bryson explained. The district court then did exactly what the Supreme Court said ***not*** to do in *Diehr* and failed to analyze the claims as a whole. Instead, the court improperly split them into pieces and looked at the physical steps separate from the BRCA gene sequences. Viewed as a whole, these claims are patentable applications of Myriad’s discovery of the BRCA sequences. The court also improperly analyzed

⁴ The district court additionally found that the balance of hardships factor “tips slightly” in favor of Ambry in view of its section 101 finding, and concluded that the public interest did not mandate either the imposition or denial of an injunction. [A100-06.]

whether the method claims unlawfully preempt use of the BRCA sequences, adopting a test—supported nowhere in the law—that the claims are impermissibly preemptive if they recite the “most widely used” method of analyzing the sequence.

On the primer pair claims, the district court erred in finding them to be ineligible products of nature, again misapplying Supreme Court precedent. The claimed primer pairs are not products of nature—they are designed and made by scientists in a laboratory and have a distinct utility from isolated DNA fragments. The district court essentially treated the claims as if they were directed to isolated DNA fragments, ignoring that they specifically recite *primers*, making them far narrower than any of the claims at issue in *AMP*. The claimed primer pairs, unlike natural DNA or isolated fragments, are useful for “priming,” or initiating the PCR process in a laboratory. The district court’s holding that *AMP* somehow implicitly suggested that DNA primers were ineligible is a vast overread of the decision and contrary to the express record before this Court and the Supreme Court.

Although the district court properly found that Myriad would be irreparably harmed in the absence of an injunction, it erred in analyzing the other factors, allowing its improper section 101 analysis to infect its finding on the balance of harms and failing to give sufficient weight to the public’s interest in protecting substantial investment in the development of life-saving technology like Myriad’s.

Because of the district court's significant errors, its judgment should be reversed, vacated, and remanded for further consideration.

ARGUMENT

I. LEGAL STANDARDS

This Court reviews a district court's denial of a preliminary injunction for abuse of discretion. *Robert Bosch LLC v. Pylon Mfg. Corp.*, 659 F.3d 1142, 1147 (Fed. Cir. 2011). When denying an injunction, "[a] district court abuses its discretion when it acts 'based upon an error of law or clearly erroneous factual findings' or commits 'a clear error of judgment.'" *Id.* A "court by definition abuses its discretion when it makes an error of law." *Koon v. United States*, 518 U.S. 81, 100 (1996); *Harris Corp. v. Ericsson Inc.*, 417 F.3d 1241, 1248 (Fed. Cir. 2005).

Courts balance four factors when considering a preliminary injunction: (1) likelihood of success on the merits of the underlying litigation, (2) whether irreparable harm is likely if the injunction is not granted, (3) the balance of hardships as between the litigants, and (4) the public interest. *See Abbott Labs. v. Sandoz, Inc.*, 544 F.3d 1341, 1362-63 (Fed. Cir. 2008). A defendant challenging validity under the first factor of the inquiry must demonstrate a "substantial question" of invalidity. *Trebo Mfg., Inc. v. Firefly Equip., L.L.C.*, -- F.3d --, 2104 WL 1377790 (Fed. Cir. 2014).

Patent eligibility under 35 U.S.C. § 101 is an issue of law that is subject to de novo review, though there may be underlying fact issues in the analysis. *Ultramercial, Inc. v. Hulu, LLC*, 722 F.3d 1335, 1342 (Fed. Cir. 2013).

II. THE DISTRICT COURT ERRED IN FINDING A SUBSTANTIAL QUESTION OF PATENT ELIGIBILITY OF THE METHOD CLAIMS

A. Representative Claims 7 and 8 Are Patent-Eligible Methods of Applying the BRCA1 DNA Sequence

The categories of patent-eligible subject matter set forth in 35 U.S.C. § 101 are broad and have always been interpreted broadly to reflect congressional intent that “the patent laws would be given a wide scope.” *Bilski v. Kappos*, 130 S. Ct. 3218, 3225 (2010). Conversely, the judicially-created exceptions to section 101—which bar patents that claim natural phenomena, laws of nature, and abstract ideas—are interpreted and applied narrowly. *Id.* at 3229; *CLS Bank Int’l v. Alice Corp.*, 717 F.3d 1269, 1277 (Fed. Cir. 2013) (“[D]anger also lies in applying the judicial exceptions too aggressively because ‘all inventions at some level embody, use, reflect, rest upon, or apply laws of nature, natural phenomena, or abstract ideas.’”).

Representative claims 7 and 8 of the ’441 patent do not fall within the narrow judicial exceptions to section 101. These claims do not purport to claim ownership of the BRCA1 gene sequence itself, nor do they repeat claim 1’s error of reciting simply an abstract comparison. Rather, they recite particular methods

of using the BRCA1 gene sequence and are patent-eligible under a proper reading of section 101.

There can be no doubt, based on decades of Supreme Court precedent, that while a natural product itself or abstract idea may not be patentable, the application of either can be. *See Bilski*, 130 S. Ct. at 3230 (“[W]hile an abstract idea, law of nature, or mathematical formula could not be patented, an **application** of a law of nature or mathematical formula to a known structure or process may well be deserving of patent protection.” (internal quotation marks omitted)); *Diehr*, 450 U.S. at 187 (“It is now commonplace that an **application** of a law of nature or mathematical formula to a known structure or process may well be deserving of patent protection.”); *Gottschalk v. Benson*, 409 U.S. 63, 67 (1972).

In *AMP*, the Supreme Court, favorably referring to Judge Bryson’s opinion concurring-in-part and dissenting-in-part in this Court’s earlier opinion in that case, again explained that applications of a natural product—specifically, genetic sequences—are patent-eligible: “Judge Bryson aptly noted that, ‘[a]s the first party with knowledge of the [BRCA1 and BRCA2] sequences, Myriad was in an excellent position to claim applications of that knowledge. Many of its unchallenged claims are limited to such applications.’” *AMP*, 133 S. Ct. at 2120 (quoting *AMP*, 689 F.3d at 1349). Notably, Judge Bryson’s analysis occurred **after**

the Supreme Court remanded *AMP* in the first instance specifically to consider the *Mayo* decision and how it applied to Myriad's patents.

In particular, Judge Bryson identified several of the then unchallenged claims as examples of patentable applications of knowledge about the BRCA sequences, including claim 21 of the '441 patent-in-suit. Representative claims 7 and 8 of the '441 patent are remarkably similar to claim 21. Each covers specific applications of the BRCA1 gene that involve comparing a human BRCA1 sample with a wild-type BRCA1 sequence using identified laboratory techniques and synthetic probes and primers. Indeed, claims 7 and 21, using nearly identical language, both require the very same steps of hybridizing a BRCA1 gene probe and detecting the presence of a hybridization product, which indicates the presence of a BRCA1 allele in a patient. Claim 8 contains different limitations of amplifying and sequencing the BRCA1 gene, but these steps are similarly concrete and physical. These types of specific, limited applications of BRCA1 knowledge are patentable under section 101.

The district court never addressed these similarities, despite the fact that Myriad raised them both in briefing and at the hearings. [A9213; A8865-66; A9789; A9796-97.] Instead, the district court, without actually discussing or analyzing any of the individual asserted method claims, grouped them all together and stated that they have "striking initial similarities" to the claims that this Court

invalidated in *AMP*. [A90.] It was legal error to group claims 7 and 8 with the invalidated claims and ignore that these claims contain precisely the kind of limitations that Judge Bryson, and other precedent, indicated makes them patent-eligible.

The method claims invalidated in *AMP* were found patent ineligible because they covered only “abstract mental processes”—specifically, they recited abstract methods for comparing or analyzing gene sequences without any physical steps. [E.g., A348 (claim 1); A542 (claims 1 and 2).] For example, this Court held that claim 1 of the ’441 patent did not recite eligible subject matter because the only step in the claimed method was “comparing” one BRCA1 sequence with another BRCA1 sequence, which “recites nothing more than the abstract mental steps necessary to compare two nucleotide sequences.” *AMP*, 689 F.3d at 1334-45. This Court went on to explain that “the application of a formula or abstract idea in a process may describe patent eligible subject matter, [but] Myriad’s claims [at issue in that case] **do not apply the step** of comparing two nucleotide sequences in a process” because the claims include no physical steps, like “sequencing the BRCA DNA molecule.” *Id.* (emphasis added).

Here, in contrast, claims 7 and 8 of the ’441 patent unquestionably **apply** the abstract step of comparing the nucleotide sequences by requiring physical limitations. Claims 7 and 8 expressly recite specific physical steps, namely

“hybridizing” and “detecting” (claim 7) and “amplifying” and “sequencing” (claim 8) using probes or primers. The district court’s comparison to the *AMP* method claims invalidated as directed to abstract ideas betrays a fundamental misunderstanding of the judicially-created exception to section 101. Claims 7 and 8 recite patentable applications of knowledge about the BRCA1 sequence, and the district court’s erroneous comparison was legal error.

B. Representative Claims 7 and 8, Properly Viewed as a Whole, Contain an “Inventive Concept” Under *Mayo*

Compounding its improper comparison error, the district court then attempted to apply *Mayo* to Myriad’s method claims, finding that “the only ‘inventive concepts’ in [Myriad’s] Method Claims are the patent-ineligible naturally-occurring BRCA1 and BRCA2 sequences. The claims contain no other new process for designing or using probes, primers, or arrays beyond the use of BRCA1 or BRCA2 in those processes.” [A93.] This is a serious misapplication of the “inventive concept” that the Supreme Court discussed in *Mayo* and is contrary to other Supreme Court precedent.

In *Mayo*, the Supreme Court explained that its section 101 precedent “insist[s] that a process that focuses upon the use of a natural law also contain other elements or a combination of elements, sometimes referred to as an ‘inventive concept,’ sufficient to ensure that the patent in practice amounts to significantly more than a patent upon the natural law itself.” 132 S. Ct. at 1294.

The Court then explained that the steps in Prometheus’s claims did not satisfy those conditions because they “involve well-understood, routine, conventional activity previously engaged in by researchers in the field.” *Id.*

But the “inventive concept” discussion in *Mayo* is not an invitation, as the district court appeared to believe, to carve the claims into pieces—the natural law on one side and the remaining steps on the other—and evaluate the remaining steps wholly apart from the natural law to determine if they are “conventional” or “routine.” In breaking up the claims the way it did, the district court violated the long-standing rule from *Diehr* that “[i]n determining the eligibility of respondents’ claimed process for patent protection under § 101, their claims must be considered **as a whole**. It is inappropriate to dissect the claims into old and new elements and then to ignore the presence of the old elements in the analysis.” 450 U.S. at 189. Indeed, this portion of *Diehr* explicitly overruled contrary authority in *Flook*, which considered the unpatentable subject matter as part of the prior art, and which the district court mistakenly cited as support for its analysis. *Id.* at 189 n.12. [A91-93.]

The district court here did exactly what the Supreme Court in *Diehr* said was wrong—ignored the BRCA DNA sequences and simply looked to the other steps in the claims in the abstract to determine if they are “routine” or “conventional.” *Mayo* does not instruct the court to do this type of analysis. Rather, *Mayo* explains

that the claims at issue there were invalid because they merely “inform a relevant audience about certain laws of nature; any additional steps consist of well-understood, routine, conventional activity already engaged in by the scientific community; and those steps, *when viewed as a whole*, add nothing significant beyond the sum of their parts taken separately.” 132 S. Ct at 1298 (emphasis added).

When representative claims 7 and 8 of the '441 patent are viewed *as a whole*, without putting aside the BRCA1 DNA sequence, it is clear that they recite more than “routine,” “conventional” steps previously used in the art at the time of Myriad’s patent applications. Indeed, the BRCA1 sequence was unknown at the time of the inventions, and the hybridization/detection and amplification/sequencing steps recited in claims 7 and 8, respectively, use probes and primers that Myriad specially designed for the BRCA1 gene *after* it was discovered. [A9144-46.] These steps were not previously performed in the art and, rather than being routine or conventional, were unknown. While Myriad certainly did not invent probes and primers, that does not mean Myriad may not apply them with the BRCA1 gene and satisfy section 101. *See Diehr*, 450 U.S. at 187 (“It is now commonplace that an application of a law of nature or mathematical formula to a *known* structure or process may well be deserving of patent protection.”) (emphasis added).

The invalidated *Mayo* claims, by contrast, covered a method whose steps had all been identically done in the past for the identical purpose, combined in the identical way—the method required administration of a previously-used drug for a disease it had long been used to treat, and then measured the same biomarker for that drug in the same way it had been measured for decades. *Mayo*, 132 S. Ct. at 1297-98. The *Mayo* claims then merely appended the natural correlation to these identical steps that had been around for decades, and, by doing so, attempted to monopolize activity that doctors had already been performing through the correlation. Thus, the claim was effectively to the natural correlation itself, as it barred doctors from considering it in the context of the method they had already been performing.

By contrast, claims 7 and 8 set forth a new method of processing a patient sample to diagnose cancer risk through the use of new probes and primers and a new gene sequence. Before Myriad's discovery of the BRCA1 gene, it would have been impossible to employ the claimed hybridizing, sequencing, and detecting steps because each of those steps depends on the knowledge of the BRCA1 sequence.

In denying Myriad's motion, the district court erroneously believed that Myriad's discovery of the BRCA1 sequence was irrelevant to the inventive step analysis. [A95 ("[Myriad] ask[s] the court to find that obtaining knowledge of the

naturally occurring BRCA1 and BRCA2 sequences is somehow an inventive step sufficient to render the Method Claims patent eligible.”).] While the district court remarkably faulted Myriad for allegedly citing no authority to support its arguments in this regard [A99], the reality is that there is no support for the district court’s analysis, and it again reflects the district court’s failure to look at the claims as a whole.

The suggestion that a supposedly natural sequence can play no role in patent eligibility is inconsistent with the Supreme Court’s holding in *AMP* itself. In the case, the Court found claims to cDNA eligible despite the fact that it was well-known how to synthesize cDNA once a natural sequence was discovered. *AMP*, 133 S. Ct. at 2119. Although those claims were product claims and not method claims, it is plain from the Supreme Court’s opinion that an appropriate method claim relating to the use of cDNA would have been similarly valid, despite the fact that the creation of cDNA involves laboratory processes that, considered in the abstract, were surely “conventional” at the time of Myriad’s patent. Similarly, claims 7 and 8 are also patent-eligible because they use the BRCA1 sequence as part of a method that applies that sequence in a useful way that employs new probes and primers that are both concrete and non-conventional. when considering the claims as a whole. The district court’s “inventive step” analysis was legal error, and an abuse of discretion.

C. **Representative Claims 7 and 8 Do Not Unlawfully Preempt the BRCA1 DNA Sequence**

The district court also erroneously found that the method claims were likely invalid because they allegedly preempt further use and study of the BRCA genes. Once again, the district court's application of the law was incorrect.

As an initial matter, preemption has never been adopted as the test for section 101. Despite extensive briefing in *Mayo* primarily addressing this argument, the Supreme Court's opinion does not adopt preemption as the test. *Mayo*, 132 S. Ct. at 1302 ("The presence here of the basic underlying concern that these patents tie up too much future use of laws of nature ***simply reinforces*** our conclusion that the processes described in the patents are not patent eligible, while eliminating any temptation to depart from case law precedent.") (emphasis added). While certainly the Supreme Court has stated that preemption is a consideration in the section 101 analysis, its failure to adopt it as the test, despite many opportunities to do so, is telling.

But even putting that aside, the preemption relevant for section 101 only potentially occurs where the claim covers every practical application of the unpatentable subject matter. Speaking in terms of abstract ideas, this Court in *Ultramercial* explained that "[i]t is not the breadth or narrowness of the abstract idea that is relevant, but whether the claim covers every practical application of that abstract idea." 722 F.3d at 1346. Restating that same standard for natural

products or natural laws, the question is whether the claim preempts every practical application of that natural product or natural law.

The ineligible claims in *Mayo* are a good example of impermissible preemption. Those claims foreclosed any party, not only from using the correlation in the claimed method, but even from trying to develop a new way of utilizing the correlation. The claimed “administering” and “determining” steps were necessary prerequisites to make any use of the correlation at all—one could not consider the correlation without administering the drug and measuring the metabolites. *Mayo*, 132 S. Ct. at 1298 (“Anyone who wants to make use of these laws must first administer a thiopurine drug and measure the resulting metabolite concentrations, and so the combination amounts to nothing significantly more than an instruction to doctors to apply the applicable laws when treating their patients.”). The claim thus prevented others from using the correlation for any purpose at all, even from further developing it in other ways.

In this fashion, the precatory steps in *Mayo* were “inherent” in the correlation, as this Court has described in *Ultramercial*, where it noted that “the Supreme Court’s reference to ‘inventiveness’ in *Prometheus* can be read as shorthand for its inquiry into whether implementing the abstract idea in the context of the claimed invention inherently requires the recited steps.” 722 F.3d at 1348. Stated in the context of a natural phenomenon, the inquiry is whether the claimed

application of the natural phenomenon necessarily requires the recited steps. The claims in *Mayo* unlawfully preempted the natural law because the claimed application of the law ***necessarily*** required the recited steps of administering a drug and determining metabolite levels. *Mayo*, 132 S. Ct. at 1299-1300 (“And since they are steps that must be taken in order to apply the law in question, the effect is simply to tell doctors to apply the law somehow when treating their patients.”).

Here, that is simply not the case. The recited probe hybridizing and detection steps in claim 7, and primer amplification and sequencing steps in claim 8, are not inherent in any use of the BRCA1 gene—they are not required to isolate the gene for study, and they do not prohibit studying it in other ways. The claims thus do ***not*** preempt all practical uses of the naturally occurring BRCA1 DNA. Even now, there are technologies available that utilize unenriched, natural DNA extracted from a patient’s cells and do not require production of synthetic DNA for subsequent analysis that are not implicated by Myriad’s patents.⁵ Claims 7 and 8 do not prevent others from currently using the BRCA1 gene in these other ways.

⁵ Those technologies include gene expression profiles, untargeted single-molecule sequencing, and protein truncation testing. See, e.g., <https://www.nanoporetech.com/technology/analytes-and-applications-dna-rna-proteins/dna-sequencing-applications>; <https://www.nanoporetech.com/technology/analytes-and-applications-dna-rna-proteins/dna-an-introduction-to-nanopore-sequencing> (last visited Apr. 17, 2014). [A9152-54; A6652-53.]

But more critical to the preemption analysis, the claims certainly do not prevent others from developing new ways to use and evaluate the BRCA1 gene in the future—for example, methods that do not involve primer-based amplification or probe-based hybridization, which is the key concern in the section 101 preemption analysis. *See Mayo*, 132 S. Ct. at 1302 (explaining that claims were invalid because they “cover[ed] all processes that make use of the correlations after measuring metabolites, including *later developed* processes that measure metabolite levels in new ways”) (emphasis added); *O’Reilly v. Morse*, 56 U.S. 62, 113 (1854) (expressing concern that a “future inventor, in the onward march of science, may discover” new means of applying the natural law that would still be covered by the claim); *Benson*, 409 U.S. at 68 (explaining that the “claim is so abstract and sweeping as to cover both known and unknown uses” of the abstract idea).

In this context, claims 7 and 8 compare favorably to claims in past cases that have been found patent-eligible. The claims in *Morse* provide an illustration of the contrast between ineligible claims that are preemptive and eligible claims that are properly limited to an application of a natural phenomenon or abstract idea. *Morse*’s broad claims to the use of electricity in any way to communicate “intelligible characters” were found invalid. But the narrower claims that recited particular methods of using “the motive power of magnetism” “as a means for

operating or giving motion to machinery, which may be used to imprint signals upon paper or other suitable material, or to produce sounds in any desired manner, for the purpose of telegraphic communication at any distance” were patent-eligible and not overly-preemptive. *Morse*, 56 U.S. at 120. Claims 7 and 8 are far more similar to the eligible *Morse* claims because they do not simply recite the BRCA1 sequence itself or any way of determining that sequence, but rather recite specific steps to study the sequence and make comparisons between the normal sequence and a patient’s DNA.

Following *Morse*, Alexander Graham Bell’s claims for using electrical current in a closed circuit, under certain conditions, for the transmission of vocal or other sounds were found to be patentable. The Court explained that the claim “[wa]s not for the use of a current of electricity in its natural state as it comes from the battery,” but rather to a specific method. *The Telephone Cases*, 8 S. Ct. 778, 782 (1888). The Court later explained that “Bell’s claim, in other words, was not one for all telephone use of electricity.” *Benson*, 409 U.S. at 69. Similarly, claims 7 and 8 are not for all uses of the BRCA1 sequence—they do not simply recite “apply the BRCA1 sequence” or “diagnose a patient using the BRCA1 sequence,” but recite specific and limited steps like the eligible claims in *Morse* and *The Telephone Cases*.

Claims 7 and 8 also do not fall into the category of ineligible claims in which the “purported limitations provide no real direction, cover all possible ways to achieve the provided result, or are overly generalized.” *CLS Bank*, 717 F.3d at 1301; *see also Mayo*, 132 S. Ct. at 1300 (“[S]imply appending conventional steps, specified at a high level of generality, to laws of nature, natural phenomena, and abstract ideas cannot make those laws, phenomena, and ideas patentable.”). These ineligible claims share the common feature of putting no limit on the process of using the natural phenomenon or abstract idea. In *Mayo*, “the ‘determining’ step tells the doctor to determine the level of the relevant metabolites in the blood, through whatever process the doctor or the laboratory wishes to use.” 132 S. Ct. at 1297. Similarly, the patent in *Flook* did not “contain any disclosure relating to the chemical processes at work, the monitoring of process variables, or the means of setting off an alarm or adjusting an alarm system. All that it provides is a formula for computing an updated alarm limit.” *Parker v. Flook*, 437 U.S. 584, 586 (1978). Effectively, the claims recited the formula itself.

The claims here do not suffer from that problem. They do not simply recite a general limitation such as “determining” a DNA sequence. Rather, they recite specific steps of primer-based amplification, probe-hybridization, and detection through sequencing. Because of these specific physical steps, the claims do not unlawfully preempt anything under section 101. *See Diehr*, 450 U.S. at 192-93.

Rather than carefully apply preemption consistent with the above precedent, the district court instead crafted its own “most widely means” standard, *i.e.*, if the claims preempt the “most widely used means” of using the gene, then they are impermissible. [A95.] There is no support for this in the law, and, indeed, would be a standard constantly in flux. What might be most widely used at the beginning of a patent term—and, under the district court’s reasoning, impermissibly preemptive—might be ancient technology at the end of the term—and thus, no longer impermissibly preemptive. This is not the preemption concept as applied by this Court and the Supreme Court.

As if to prove this point, the district court relied for its preemption analysis primarily on the *Sequenom* case currently on appeal to this Court and its introduction of a “commercially viable alternative” test for impermissible preemption. While Myriad does not agree that *Sequenom* was correctly decided, the claims that were invalidated in that case are readily distinguishable from the claims at issue here. *Sequenom* involves the permissible scope of patenting methods for detecting a known product of nature—fetal DNA—that was found to exist in a new location—a pregnant mother’s serum and plasma. In exploiting their discovery, the *Sequenom* inventors were thus able to use a known natural product, off-the-shelf reagents, and previously-designed primers and probes, albeit in important new ways that ought to be patentable. [A9971-73 (1:12-15, 31-32;

5:7-14).] Essentially, the *Sequenom* inventors changed the source from which they obtained the fetal DNA, but everything else was the same as in the past.

By contrast, claims 7 and 8 at issue here deal with the application of previously-unknown DNA sequences for which Myriad had to design a new PCR reaction using its own new primers and probes, which had never been previously studied and assessed before Myriad's work. The district court's comparison to *Sequenom* was flawed, and using the case as support for the district court's "most widely used means" standard—which otherwise is not found in the law—was additional error.

Claims 7 and 8 of the '441 patent permissibly apply the inventors' discovery of the BRCA1 gene and do not unlawfully preempt use and study of that gene. The district court's contrary determination was a clear abuse of discretion.⁶

⁶ Even the PTO's newly-issued guidelines for determining subject matter eligibility under section 101 suggest strongly that claims 7 and 8 are not impermissibly preemptive. While, like many, Myriad believes the guidelines reflect an overly-restrictive interpretation of section 101, they provide as an example of a patent eligible method claim a method for amplifying a natural DNA sequence using Taq polymerase and thermal cycling, *i.e.*, PCR. GUIDANCE FOR DETERMINING SUBJECT MATTER ELIGIBILITY OF CLAIMS RECITING OR INVOLVING LAWS OF NATURE, NATURAL PHENOMENA & NATURAL PRODUCTS 11, *available at* http://www.uspto.gov/patents/law/exam/myriad-mayo_guidance.pdf. The guidelines reason that these limitations "are meaningful, because others are not substantially foreclosed from using the natural products in other ways, e.g., others may use the target DNA in other methods or compositions." *Id.* at 13.

III. THE DISTRICT COURT ERRED IN FINDING A SUBSTANTIAL QUESTION OF PATENT ELIGIBILITY OF THE PRIMER PAIR CLAIMS

A. A Composition of Matter Is Patent Eligible under Section 101 and Supreme Court Precedent Unless it is a “Product of Nature”—*i.e.*, Both its Structure and Utility Are as in Nature, Unaltered by Man

For claims directed to compositions of matter, the patent eligibility inquiry focuses on whether or not the composition is considered a “product of nature.” While the Supreme Court’s case law on this topic is limited, a careful review of it demonstrates the importance of identifying not just whether the composition sought to be patented is found unaltered in nature, but also whether man has given to it a different utility from nature—*i.e.*, has man applied it to a new end. Thus, in *Funk Brothers*, a mixed culture of naturally occurring strains of bacteria was found to be an unpatentable product of nature, even though the combination itself was not found in nature. Because the bacteria were unchanged in the combination, either in identity—*i.e.*, they were the same bacteria found in nature—or in utility—*i.e.*, when combined, their behavior as bacteria was unaltered by man, section 101 barred the combination. 333 U.S. at 131-32 (explaining that the combination had “no new bacteria, no change in the six species of bacteria, and no enlargement of the range of their utility. Each species has the same effect it always had. The bacteria perform in their natural way.”). Conversely, a bacterium created by combining an existing species of bacteria with naturally occurring plasmids to

create a new bacteria that ate oil “plainly qualifie[d] as patentable subject matter.” *Diamond v. Chakrabarty*, 447 U.S. 303, 309 (1980). The combination did not exist in nature, and also had a different utility—consuming oil—that had been engineered by man. According to the Supreme Court, it had a “distinctive name, character and use.” *Id.* at 309-10.

This line between products of nature that are not patentable, on the one hand, and creations of man, on the other, is well illustrated in *AMP*, in the Supreme Court’s twin findings that “isolated DNA” claims that encompassed natural DNA were not patentable under section 101, but those limited to cDNA were. *AMP*, 133 S. Ct. at 2111. The “isolated DNA” claims at issue there encompassed all forms of “isolated DNA” that had a BRCA1 sequence, regardless of source, including DNA “isolated” only “by breaking the covalent bonds that connect the DNA to the rest of the individual’s genome.” *Id.* at 2113.

As to this broadly claimed form of “isolated DNA,” even though it does not exist in nature, isolating it did not either change the DNA structure of the genetic unit itself or give the DNA any distinctive utility sufficient enough to render it patentable. On structure, the Supreme Court stated: “It is undisputed that Myriad did not create or alter any of the genetic information encoded in the BRCA1 and BRCA2 genes. The location and order of nucleotides are as Myriad found them.” *AMP*, 133 S. Ct. at 2116. Similarly, on utility: “Nor are Myriad’s claims saved by

the fact that isolating DNA from the human genome severs chemical bonds and thereby creates a non-naturally occurring molecule. Myriad’s claims are simply not expressed in terms of chemical composition, nor do they rely in any way on the chemical changes that result from the isolation of a particular section of DNA.” *Id.* at 2118.

As to the cDNA, it unquestionably did not exist in nature, even though its sequence was dictated by nature—the mRNA from which it is synthesized. While the sequence of cDNA may be dictated by nature, “the lab technician unquestionably creates something new when cDNA is made.” *Id.* at 2119. Because it was “unquestionably” new, the Supreme Court did not comment on its utility, but cDNAs plainly have a new utility. [See A6806.]

From this backdrop, the error in the district court’s conclusion that Myriad’s primer pair claims were unpatentable products of nature is apparent. Representative claims 16 and 17 of the ’282 patent are directed to “a pair of single-stranded DNA primers for determination of a nucleotide sequence of a BRCA1 gene by a polymerase chain reaction,” where the use of the primers in PCR results in the synthesis of DNA having at least part of the sequence of the BRCA1 gene (claim 16) or the sequence of the BRCA1 cDNA (claim 17). [A242; A6816.] As demonstrated below, while they rely on the wondrous natural properties of DNA, these “primer pair” claims are limited to compositions of matter made by men,

with a utility engineered by man that is distinct from any product of nature—*i.e.*, use in PCR to initiate DNA synthesis.

B. The Primer Pair Claims Are Not a Product of Nature

1. The Primer Pair Claims Are Designed and Made by Scientists, Not Snipped from Nature

Unlike a new mineral discovered in the earth, a leaf snapped from a tree, or a natural DNA segment isolated from a human cell, the claimed primer pairs—which are short, single-stranded DNA segments used in PCR for amplifying and subsequently sequencing all or part of the BRCA1 gene—do not exist in nature. There is no dispute between the parties that there are no single-stranded DNAs that can act as primers present in a human cell, let alone a pair of such DNAs.

Instead, like the genetically engineered bacterium in *Chakrabarty* and the cDNA in *AMP*, the claimed primer pairs are designed and made by scientists in a laboratory. As the patent specification teaches, the claimed primer pairs are designed to prime (or initiate) the amplification of a specific segment of the BRCA1 gene (or a portion of the sequence of BRCA1 cDNA, as in the case of dependent claim 17) in PCR, and the sequences of the primer pairs are chosen by the scientist to uniquely and separately hybridize to two regions of DNA flanking the target segment. [A172-73 (16:23-35; 17:23-27); *see also* A6806; A1954-57; A1960-62.] As the district court acknowledged, “[s]cientists create primers. In so doing, they consider primer size and other aspects, such as the exact portion of the

DNA segment targeted.” [A14; *see also* A13 (“A primer is a short, ***synthetic***, single-stranded DNA molecule that binds specifically to an intended target nucleotide sequence.”) (emphasis added).]

The fact that the claims are directed to primer *pairs* moves them even further from the realm of products of nature. A scientist does not just design one primer and stop there. Rather, the scientist designs a coordinated pair that, working in conjunction, primes a non-natural chemical reaction that does not occur in the body, *i.e.*, PCR. The use of a *pair* of primers, as recited in the claims, is critical. Unless the two coordinated primers are used together, they cannot perform their function. This is the work of scientists in the laboratory, not the work of nature.

[A1954-57; A1960-62; A6806.]

That the nucleotide sequences of a pair of primers may, at some level, be said to be “dictated by the sequence of the gene that the scientist is trying to amplify”⁷ does not negate the fact that the primer pairs are the creation of the scientist, not nature. [See A6338-39; *see also* A14.] “The rule against patents on naturally occurring things is not without limits,” for “all inventions at some level embody, use, reflect, rest upon, or apply laws of nature, natural phenomena, or

⁷ Even here, it is the scientist who designs the primers based on the segment of the gene that he or she wishes to amplify. Thus, unlike cDNA, where the sequence is directly dictated by the RNA, the sequence of a primer depends on exactly what the scientist is trying to amplify. [A1954-57; A1960-62; A6805; A7614.]

abstract ideas,” and “too broad an interpretation of this exclusionary principle could eviscerate patent law.” *AMP*, 133 S. Ct. at 2116 (quoting *Mayo*, 132 S. Ct. at 1293). The same “dictated by nature” argument was expressly rejected by the Supreme Court in *AMP* in finding cDNA patent eligible. As the Court noted, although the nucleotide sequence of cDNA may be dictated by nature, “the lab technician unquestionably creates something new when cDNA is made.” *Id.* at 2119. Likewise, here, the scientist unquestionably creates something new when the primer pairs are made.

In this regard, the primer pair claims do not cover isolated natural DNA fragments that share the same sequence. Natural DNA fragments are just that—fragments of DNA. They can be used for whatever use one can make of them. Conversely, primers are just that—primers. Claims to them only cover use of the claimed material in PCR, an entirely separate, man-made utility, as demonstrated below. In short, while a DNA fragment might be capable of being utilized by a scientist as a single primer, a claim to a pair of primers ***functioning together*** for use in PCR does not claim the unpatentable DNA fragment alone, divorced from the pair’s application in PCR.

2. The Claimed Primer Pairs Have Utility Distinct from Natural DNA or Fragments Thereof

As the name suggests, the function of a primer is to “prime,” *i.e.* to serve as a starting material for a DNA polymerization process. [See A1954-55; A7617-18.]

During PCR, each single-stranded DNA primer of a primer pair hybridizes (or binds) to a specific portion of the template DNA; together, they define the boundaries of the DNA segment in interest. [See A1954-57; A6806-07.] DNA synthesis is initiated with the DNA polymerase adding a nucleotide complementary to the template DNA to the free 3'-end of each of the primers. DNA polymerase then continues to add nucleotides to the free 3'-end of the elongating strand in an order complementary to the template DNA, resulting in a new strand of DNA. [See A6807.]

Naturally occurring genomic DNA, on the other hand, is a molecule that stores the biological information used in the development and functioning of all known living organisms including humans, and contains the code and instructions necessary for making virtually all proteins in a cell. [See A7-8.] Natural DNA does not and cannot serve the “priming” function during DNA replication in a human body. During DNA replication inside a cell, as the double-stranded DNA unwinds, each of the two single strands becomes available as a **template**, not as a **primer**, for replication. [A6309-10; A7617.] Specifically, an enzyme called a “primase” synthesizes a single short segment of RNA using one of the DNA strands to be replicated as a template. During its synthesis, the “RNA primer” is created one nucleotide at a time, and each nucleotide is annealed or attached directly to the DNA strand being replicated one nucleotide at a time. In other

words, the primer is never created as a separate, single-stranded molecule, but instead is created directly on the DNA strand. Once this “RNA primer” is formed directly on the DNA template, the cell’s DNA polymerase uses that primer as the starting point for DNA elongation. [A7616-17; A7635-36; *see also* A9-10.]

Accordingly, the DNA replication process in cells is distinct from the laboratory PCR process because: (1) DNA replication uses “RNA” primers that never exist as a separate single-strand because they are synthesized one nucleotide at a time directly onto (attached to) the DNA strand to be replicated, in contrast to using DNA primers that are single-stranded; (2) only one RNA primer is used for each part of the DNA strand replicated rather than a matched “pair” or primers that allows for synthesis of DNA only between the matched pair of primers; (3) PCR involves the use of fragmented pieces of input DNA as templates in contrast to using the unbroken continuous DNA as it exists in the genome during DNA replication; and (4) PCR uses many cycles (20-30 cycles) of heating and cooling to completely separate or “denature” the double-stranded DNA into single strands in each cycle, and this cyclical heating and cooling process is not used in the natural DNA replication process. [A7616-17; A7635-36; A6309-10; A6318-20; A6336-37; A6281-82; A6806-07.]

The district court’s failure to give weight to this distinct utility for primers from natural DNA or its fragments appears to be based largely on the fact that a

DNA primer has a sequence identical to a fragment of natural DNA. [A83-85.] While the claimed DNA primer pairs may share the same sequence with two fragments of natural DNA, they have an entirely different function. As such, the claimed DNA primer pairs are not the same as a fragment of natural DNA, even though they share the same sequence, but rather are a patent eligible *application* of fragments of natural DNA. The claims do not cover the natural DNA fragments with which they share a sequence, but instead utilize knowledge of the sequence in a patent-eligible way, to aid in the conduct of PCR. The district court erred by ignoring the differences in utility and focusing solely on the identity of the nucleotide sequences.⁸

The district court also placed undue emphasis on the fact that the claimed DNA primers carry out their distinct priming function by hybridizing to the target DNA sequence “according to Watson-Crick pairing.” [A85.] But “all inventions at some level embody, use, reflect, rest upon, or apply laws of nature, natural phenomena, or abstract ideas.” *Mayo*, 132 S. Ct. at 1293. A composition of matter

⁸ In this regard, the primer pair claims at issue here are much narrower than the hypothetical primer pair claim in the PTO Guidance Memorandum. *See supra* at n.6. There, setting aside the fact that the PTO applied an unduly broad interpretation of the Supreme Court’s decisions, the claimed “pair of primers” are not limited to any specific application; the only limitation is that the primers have the sequences of SEQ ID No:1 and SEQ ID No:2, respectively, which, in PTO’s hypothetical, are “naturally occurring DNA sequences found on a human chromosome.” *Id.* at 11-12.

is not unpatentable simply because its utility relies on a law of nature. *Id.*; *AMP*, 133 S. Ct. at 2116. Indeed, the Supreme Court already ruled that BRCA1 cDNA—which uses the same Watson-Crick pairing principal—is patentable. *AMP*, 133 S. Ct. at 2119. Rather, it is only if the composition’s utility is the same as in nature, unaltered by man, does the reliance on natural properties for utility become impermissible.

Thus, unlike the combination of bacteria in *Funk Brothers*, the claimed pair of primers is not a mere mixture of two independent natural products with “no enlargement of the range of their utility.” 333 U.S. at 131. Instead, the claims are to two interdependent molecules carefully designed by a scientist to work in a coordinated fashion to catalyze a non-natural PCR reaction. The pair of primers does not “perform in their natural way” to “serve the ends nature originally provided.” *Id.* Rather, like the genetically engineered bacterium in *Chakrabarty*, the claimed primer pairs have a utility “possessed by no naturally occurring” DNA; the primer pair claims here are therefore “not to a hitherto unknown natural phenomenon, but to a non-naturally occurring manufacture or composition of matter—a product of human ingenuity having a distinctive name, character and use.” *Chakrabarty*, 447 U.S. at 309-10. The district court was wrong to find otherwise.

C. The District Court Vastly Overread *AMP* to Include Primers and Probes as Patent Ineligible

In finding the primer pair claims likely invalid, the district court misinterpreted the Supreme Court’s decision in *AMP* in a variety of ways. As an initial matter, the district court read the decision as standing for the rule that “isolated DNA segments were patent ineligible as long as they *reflected* naturally occurring BRCA1 and BRCA2 sequences.” [A75 (emphasis added).] This is unduly broad. Rather, as the district court stated at the outset of its opinion, the Supreme Court’s holding in *AMP* was narrow: “We merely hold that genes and the information they encode are not patent eligible under § 101 simply because they have been isolated from the surrounding genetic material.” *AMP*, 133 S. Ct. at 2120. The Court expressly rejected the petitioners’ “dictated by nature” argument and found BRCA cDNA patent eligible, even though they too “reflect” naturally occurring BRCA sequences. *Id.* at 2119.

The Supreme Court’s comment that “cDNA is not a ‘product of nature’ and is patent eligible under section 101, except insofar as very short series of DNA may have no intervening introns to remove when creating cDNA,” *see AMP*, 133 S. Ct. at 2119, does not change matters. The district court failed to appreciate the context of the Court’s statement in *AMP*, which was made, and must be interpreted, in the context of the “isolated DNA” claims before the Court. The Court’s reference to “cDNA” was shorthand for its inquiry into whether an

“isolated” DNA having the sequence of a particular cDNA is patent eligible. Claim 2 of the ’282 patent, for example, is directed to an “isolated DNA . . . [having] the nucleotide sequence set forth in SEQ ID NO:1.” *Id.* at 2113. While claim 2 does not expressly recite a “cDNA,” the Court found that for all intent and purposes, claim 2 was limited to BRCA1 cDNA in view of the recitation of “SEQ ID NO:1,” which is the sequence of the cDNA. *See id.*; *see also AMP*, Oral Argument Tr. (Apr. 15, 2013) at 22:13-23.

On the other hand, had the cDNA sequence been “very short” with “no intervening introns to remove,” an “isolated DNA” having that short cDNA sequence would not only have covered the cDNA but also genomic DNA with the same sequence, and therefore may have been patent ineligible under section 101. In other words, in the context of the claims before the Court, an “isolated DNA having the nucleotide sequence” of a very short cDNA may be patent ineligible not because the cDNA may be patent ineligible even though it is man-made and not a product of nature, but because the “isolated DNA” in that case may encompass genomic DNA—*i.e.*, that which exists in nature.

But more than all of this, the district court was simply wrong to find, at Ambry’s repeated urging, that this Court (and by extension the Supreme Court) was actually concerned with the patentability of primers and probes, ***claimed as primers and probes***, in its prior decisions. [A77-81; A2441-43.] Although the

district court recognized that the Supreme Court never explicitly addressed the patent eligibility of DNA primers, the district court nonetheless concluded that such a holding was implicit because both *Myriad* and this Court had stated that short strands of isolated DNAs could be used as primers and probes, and thus, when claims to isolated fragments were invalidated, implicitly primers and probes were invalidated as well. [A77; A81-82.]

This betrays a fundamental misunderstanding of the basics of how a patent claim is rendered invalid. A claim is invalid under any section of the Patent Act, including section 101, so long as it encompasses anything that is invalid, in this case, any patent ineligible subject matter. *See, e.g., Titanium Metals Corp. v. Banner*, 778 F.2d 775, 780-82 (Fed. Cir. 1985) (where “a claim covers several compositions, the claim is ‘anticipated’ if one of them is in the prior art”). The Supreme Court’s holding that claim 1 of the ’282 patent was invalid under section 101 because it encompassed natural, genomic BRCA1 DNA and fragments thereof does not mean that all subject matters covered by claim 1 are patent ineligible. Indeed, claim 1 also encompassed BRCA1 cDNA, the subject matter of dependent claim 2, which the Court found patent eligible. *AMP*, 133 S. Ct. at 2113, 2119.

For the district court to assert the opposite, and to criticize *Myriad* in the process as “revers[ing] course” for asserting that primers (and probes) are patentable [A78], is an error so fundamental as to call into question the district

court's entire analysis. The Supreme Court explicitly cautioned against the overly broad interpretation adopted by the district court and stated that the *AMP* decision “*merely* h[e]ld that genes and the information they encode are not patent eligible.” *AMP*, 133 S. Ct. at 2120 (emphasis added). As discussed above, DNA primers used in PCR neither are “genes” nor “encode” any information, and claims to “primers” do not cover fragments of isolated DNA in and of themselves. They cover “primers.” Period.

Indeed, the Supreme Court's decision was silent on primers and probes because the primer and probe claims were not at issue in *AMP* and were intentionally excluded from the proceedings by AMP. As noted above, during oral argument, Justice Sotomayor questioned petitioners' counsel about this very subject: “The primers and probes stand”? *AMP*, Oral Argument Tr. (Apr. 15, 2013) at 11. The petitioners agreed: “Even if you were to rule for Petitioners, you would not have to rule concerning the use of DNA as a probe or a primer.”⁹ *Id.*

The district court was also mistaken when it observed that “Judge Bryson did not explicitly use the term ‘primer’ in relation to the non-cDNA isolated

⁹ ACLU, who brought the *AMP* suit and argued before the Supreme Court, also filed an amicus brief in this case in support of Ambry's opposition to the preliminary injunction motion. Tellingly, ACLU only challenged the patentability of the method claims and was conspicuously silent on the claims directed to pairs of primers. [See A6437-39; A6448-49; A6451-56.]

DNA,” but nonetheless asserted that Judge Bryson’s rationale would apply to primers. [A77.] In fact, Judge Bryson did use the term “primer” but came to a completely different conclusion. In finding the claims directed to short segments of DNA sequences unpatentable under section 101, Judge Bryson noted that these claims were not *limited* to primers or probes, and that Myriad could easily have claimed more narrowly to achieve the utility it attached to short segments of DNA. *AMP*, 689 F.3d at 1356. Claims narrowly drawn to DNA primers are not unpatentable simply because the broad “isolated DNA” claims are.

These errors of law surrounding the district court’s interpretation of *AMP* simply compound the court’s failure to appreciate that primer pairs are different in structure and utility than unpatentable isolated DNA. Under the correct interpretation of this Court’s and Supreme Court precedent, the primer pair claims are valid under section 101. The district court’s erroneous interpretation of this precedent was legal error, and an abuse of discretion. *Koon*, 518 U.S. at 100.

IV. The District Court Abused Its Discretion In Analyzing the Balance of Hardships and Public Interest Factors

The district court’s analysis of the balance of hardships is plagued by its erroneous finding that the asserted claims were likely invalid under section 101. Indeed, the only reason the district court found that the balance tipped in Ambry’s favor was that it believed Ambry had shown that the asserted patents were likely invalid. [A101-02.] Because the district court’s only basis for finding that the

balance of hardships favored Ambry, despite the irreparable harm the district court found Myriad would suffer absent an injunction, was wrong, Ambry is similarly situated to defendants in cases where the balance of hardships was found to favor the patentee. *See, e.g., Sanofi-Synthelabo v. Apotex, Inc.*, 470 F.3d 1368, 1383 (Fed. Cir. 2006).

On the public interest factor, the district court correctly found that Myriad “persuasively” argued that the public interest lies in upholding its patent rights, “particularly where [Myriad] . . . invested over \$500 million to improve Myriad’s testing produce; develop an extensive database of variant classifications; create a market wherein third-party payors will reimburse testing cost; and provide testing to over one million patients.” [A103-05.] Nonetheless, the district court concluded that the public interest did not favor a preliminary injunction because, in its opinion, Myriad’s patent protection of its investments allegedly hindered others’ research endeavors in this field. [A105.] Myriad should not be penalized for obtaining and using the protections of the patent laws. This Court has “long acknowledged the importance of the patent system in encouraging innovation. Indeed, the encouragement of investment-based risk is the fundamental purpose of the patent grant, and is based directly on the right to exclude.” *Sanofi-Sythelabo*, 470 F.3d at 1383-84 (internal quotation marks omitted). Just as the public interest favors granting preliminary injunctions when generic drug manufacturers seek to

enter the market in violation of pharmaceutical patents, it also favors granting preliminary injunctions in the field of personal medicine where competitors like Ambry seek to free ride off a patentee's substantial investments. *See Celsis In Vitro, Inc. v. Cellzdirect Inc.*, 664 F.3d 922, 931-32 (Fed. Cir. 2012); *Pfizer, Inc. v. Teva Pharms., USA, Inc.*, 429 F.3d 1364, 1382 (Fed. Cir. 2005).

CONCLUSION

For all the reasons set forth above, the district court's denial of a preliminary injunction was an abuse of discretion. Myriad respectfully requests that the judgment be reversed and vacated and the case remanded for further proceedings.

Dated: April 18, 2014

Respectfully submitted,

/s/ Jonathan E. Singer

Jonathan E. Singer

Deanna J. Reichel

FISH & RICHARDSON P.C.

60 South Sixth Street, Suite 3200

Minneapolis, MN 55402

Tel: (612) 335-5070

Fax: (612) 288-9696

COUNSEL FOR APPELLANT

MYRIAD GENETICS, INC.

/s/ David G. Mangum

David G. Mangum

PARSONS BEHLE & LATIMER

201 South Main Street, Suite 1800

P.O. Box 45898

Salt Lake City, UT 84111

(801) 532-1234

COUNSEL FOR APPELANTS

UNIVERSITY OF UTAH RESEARCH

FOUNDATION, THE TRUSTEES OF THE

UNIVERSITY OF PENNSYLVANIA, HSC

RESEARCH AND DEVELOPMENT

LIMITED PARTNERSHIP,

ENDORECHERCHE, INC., AND

MYRIAD GENETICS, INC

**IN THE UNITED STATES DISTRICT COURT
DISTRICT OF UTAH, CENTRAL DIVISION**

IN RE: BRCA1- AND BRCA2- BASED
HEREDITARY CANCER TEST PATENT
LITIGATION

MDL CASE No. 2:14-MD-2510

THIS DOCUMENT RELATES TO:

CASE No. 2:13-CV-00640-RJS

UNIVERSITY OF UTAH RESEARCH
FOUNDATION, *et al.*

**MEMORANDUM DECISION AND
ORDER DENYING PLAINTIFFS’
MOTION FOR PRELIMINARY
INJUNCTION**

Plaintiffs,

Judge Robert J. Shelby

vs.

AMBRY GENETICS CORPORATION,
Defendant.

On June 13, 2013, the Supreme Court issued a unanimous decision holding that “genes and the information they encode are not patent eligible simply because they have been isolated from the surrounding genetic material.” *Association for Molecular Pathology v. Myriad Genetics Corp.* (AMP), 133 S. Ct. 2107, 2120 (2013). This case arises in the aftermath of that decision.

Plaintiff Myriad Genetics, Inc. (Myriad) is recognized as the winner in the “race” to locate and sequence the BRCA1 and BRCA2 genes.¹ Myriad invested millions of dollars, including money obtained via public grants, in an effort to locate and sequence those genes in the early-to-mid-1990s. Once it did, Myriad sought and obtained related patents, some of which will begin to expire in August 2014. Myriad also developed and commercialized tests to screen people for the presence of harmful variations in these genes. Myriad launched its flagship ‘BRACAnalysis’ test in 1996 and debuted its ‘myRisk’ test in 2013. Ford Decl. at ¶¶ 1-3, 8

¹ See *Fierce Competition Marked Fervid Race For Cancer Gene*, Natalie Angier, New York Times, Sept. 20, 1994, <<http://www.nytimes.com/1994/09/20/science/fierce-competition-marked-fervid-race-for-cancer-gene.html?pagewanted=all&src=pm>>.

(Dkt. 6).² Between 1997 and 2013, Myriad's revenue from its BRCAAnalysis test steadily increased, and now totals more than \$2 billion. Kearl Decl. at 6 (Dkt. 107). Myriad earned that revenue by carefully guarding its patent rights and preventing others from providing screening tests for the BRCA1 and BRCA2 genes. From the mid-1990s, until the Supreme Court's *AMP* decision, Myriad was the lone provider of full-sequence BRCA1 and BRCA2 tests in the United States. Ford Decl. at ¶ 8.

Within days of the Supreme Court's *AMP* decision, Defendant Ambry Genetics Corporation (Defendant) announced plans to sell tests less expensive than Myriad's to screen BRCA1 and BRCA2 genes. Since then, other companies have followed suit—publicly offering such tests or announcing plans to do so.³

Soon after Defendant announced it would begin to offer BRCA1 and BRCA2 testing, Plaintiffs filed this action, complaining that Defendant's genetic testing infringes several of

² Unless otherwise noted, citations to the parties' case filings will refer to materials filed in Case No. 2:13-CV-00640, the first filed of Plaintiffs' cases in this court.

³ These companies include Gene by Gene, Ltd.; Counsyl, Inc. (Counsyl); Quest Diagnostics (Quest); GeneDx; Invitae Corporation (Invitae); and Laboratory Corporation of America Holdings (LCAH). Since the filing of the Motion for Preliminary Injunction against Defendant, Plaintiffs, including Myriad, have sued Gene by Gene, Quest, GeneDx, Invitae, and LCAH in this court. *See* Case Nos. 2:13-CV-00643 (dismissed without prejudice on Feb. 7, 2014); 2:13-CV-00967; 2:13-CV-00954; 2:13-CV-01049; and 2:13-CV-01069, respectively. Myriad was also sued in the Northern District of California by Counsyl (Case No. CV-13-04391-NC) and Invitae (Case No. 13-05495) and in the Central District of California by Quest (Case No. 8:13-CV-01587-AG-DFMx). Myriad subsequently moved the Judicial Panel on Multidistrict Litigation to transfer these California cases to the District of Utah. (MDL No. 2510.) That Panel granted Myriad's request, and the California actions have been transferred to this court as part of Case No. 2:14-MD-2510, joining this case and the cases that Plaintiffs have filed against GeneDx and Quest. Plaintiffs have not moved for injunctions against Quest, GeneDx, Invitae, or LCAH. Plaintiffs and Gene by Gene recently stipulated to dismissal without prejudice of Plaintiffs' claims and Gene by Gene's counterclaims. (Dkt Nos. 90 and 91 in Case No. 2:13-CV-00643.)

Plaintiffs' patents.⁴ Plaintiffs now move the court for a preliminary injunction enjoining Defendant's sales or offers to sell "genetic tests including a BRCA1 or BRCA2 panel" pending trial on the merits.⁵ Plaintiffs' Motion focuses on ten claims in the patents-in-suit: 1) four claims to pairs of synthetic DNA strands, called "primers"; and 2) six methods claims for analyzing BRCA1 and BRCA2 sequences. Plaintiffs argue these claims remain patent eligible after the *AMP* litigation, and that Defendant's testing infringes the patents containing these claims. Plaintiffs contend an injunction is necessary to prevent irreparable harm to their pricing structure, share of the BRCA1 and BRCA2 testing market, corporate reputation, and other exclusive benefits they might enjoy during the remainder of their patents' terms.

Defendant opposes Plaintiffs' Motion, arguing that Plaintiffs cannot show that they are likely to succeed on the merits of their infringement claims because Defendant has raised a "substantial question" concerning the subject matter eligibility of Myriad's BRCA1 and BRCA2-related patents, particularly in light of the recent *AMP* litigation. Defendant further contends there are substantial questions concerning whether: 1) its testing infringes Plaintiffs' patent claims; 2) the patents at issue are invalid because the inventions they claim were anticipated and

⁴ Plaintiffs allege Defendant is infringing: claim 6 of U.S. Patent No. 5,709,999 (the '999 Patent); claims 6, 16, and 17 of U.S. Patent No. 5,747,282 (the '282 Patent); claims 7, 8, 12, 23 and 26 of U.S. Patent No. 5,753,441 (the '441 Patent); claims 29 and 30 of U.S. Patent No. 5,837,492 (the '492 Patent); claim 4 of U.S. Patent No. 6,033,857 (the '857 Patent); claims 2, 3, and 4 of U.S. Patent No. 5,654,155 (the '155 Patent); claims 2, 3, 4, 5, 6, and 7 of U.S. Patent No. 5,750,400 (the '400 Patent); claims 32 and 33 of U.S. Patent No. 6,051,379 (the '379 Patent); claim 5 of U.S. Patent No. 6,951,721 (the '721 Patent); claims 3, 4, 5, 6, 7, 8, 11, 14, 17, 18, and 19 of U.S. Patent No. 7,250,497 (the '497 Patent); claims 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, and 18 of U.S. Patent No. 7,470,510 (the '510 Patent); claims 10, 11, 15, 16, 17, and 19 of U.S. Patent No. 7,622,258 (the '258 Patent); claims 2, 8, and 16 of U.S. Patent No. 7,838,237 (the '237 Patent); claims 2, 3, 5, 9, 10, and 12 of U.S. Patent No. 7,670,776 (the '776 Patent); and claims 2 and 7 of U.S. Patent No. 7,563,571 (the '571 Patent). Am. Compl. (Dkt. 21).

⁵ Plaintiffs' Proposed Order (Dkt. Nos. 5 and 5-1.)

obvious; and 3) the patents are invalid due to indefiniteness or lack of written description.

Defendant also argues that Plaintiffs will suffer no immediate, irreparable harm to their pricing, market share, or reputation. Finally, Defendant asserts that its business will be devastated and the public interest harmed if an injunction issues because the public would lose access to less expensive, more complete, and more innovative cancer testing.

On September 11 and 12, and October 7, 2013, the court received testimony and argument on Plaintiffs' Motion. Additionally, the parties have submitted numerous declarations from experts. Having carefully considered the relevant authorities, briefing from the parties and amici, oral argument, testimony, and the evidence, the court concludes Plaintiffs are not entitled to a preliminary injunction. The court finds that although Plaintiffs have shown they are likely to be irreparably harmed if an injunction does not issue, Defendant has raised substantial questions concerning whether any of the patent claims at issue in Plaintiffs' Motion are directed toward patent eligible subject matter under 35 U.S.C. § 101. In light of Defendant's showing, Plaintiffs are unable to establish that they are likely to succeed on the merits of their claims. Neither have Plaintiffs established that the equitable factors support issuance of the requested injunction. Having failed to satisfy their burden, Plaintiffs' Motion for Preliminary Injunction must be denied.

I. FACTUAL BACKGROUND

A. The Parties

1. Plaintiffs

Myriad is a Delaware molecular diagnostic corporation with its principal office in Salt Lake City, Utah. The University of Utah is a Utah nonprofit educational and research institution in Salt Lake City. The University of Pennsylvania is a Pennsylvania nonprofit educational and

research institution in Philadelphia. The Hospital for Sick Children is a pediatric health care and research facility located in Toronto, Ontario. Endorecherche is a Canadian medical research corporation in Ste-Foy, Quebec. Myriad owns the following patents-in-suit: the '155, '400, '379, '721, '497, '510, '258, '237, '776, and '571 Patents. *See supra* note 4 (listing complete numbers for patents-in-suit). Myriad is the exclusive licensee of the '999, '282, '441, '492, and '857 Patents. The University of Utah is the owner or co-owner of three patents at issue in this case, the '999, '282, and '441 Patents. The University of Utah, University of Pennsylvania, the Hospital for Sick Children, and Endorecherche are the co-owners of the '857 and '492 Patents. The University of Utah Research Foundation, also a Plaintiff, has received from Myriad over \$40 million in royalties under some of the patents at issue in this case over the past two decades. Pershing Decl. at ¶ 4 (Dkt. 112).

2. Defendant

Defendant is a clinical diagnostic and genomic services company in Aliso Viejo, California. In the hours after the Supreme Court issued its *AMP* decision, Defendant announced that it would begin offering a number of its own tests that include BRCA1 and BRCA2 screening. Defendant now offers a menu of at least six tests that include screening for BRCA1 and BRCA2: a combined BRCA1/BRCA2 test, BRCAPlus, BreastNext, PancNext, Ova Next, and CancerNext. Chao Decl. at ¶ 16, Exhs. B-G (Dkt. 56). Defendant's BRCA1/BRCA2 test is available for \$2,200—substantially less than the price for comparable testing offered by Myriad. *Id.*

3. Amici

The court permitted the filing of a joint Amicus Curiae brief in support of Defendant's Opposition to Plaintiffs' Motion for Preliminary Injunction. (Dkt. 79.) Amici are the American

Civil Liberties Union (ACLU) and ACLU of Utah Foundation, Inc. (ACLU Utah), Public Patent Foundation (PUBPAT), Association for Molecular Pathology (AMP), Breast Cancer Action (BCAction), and the AARP. The ACLU and PUBPAT represented the individual and organizational plaintiffs in the *AMP* litigation, including two of the amici here, *AMP* and BCAction. AARP also filed amicus briefs in the *AMP* litigation.

The ACLU describes itself as a “nationwide, nonprofit, nonpartisan organization with over 500,000 members” with the stated goal of protecting rights protected under the Constitution. *Id.* at 4. ACLU Utah is a regional affiliate of the ACLU. *Id.* PUBPAT is a not-for-profit legal services organization affiliated with the Benjamin N. Cardozo School of Law and is concerned with patent policy issues. *Id.* at 4-5. *AMP* is “an international not-for-profit professional association representing over 2,000 physicians, doctoral scientists and medical technologists who perform laboratory testing based on knowledge derived from molecular biology, genetics and genomics.” *Id.* at 5. *AMP* claims an interest in this matter because, in its view, the issues in this case will impact “the provision of and innovation in genetic testing.” *Id.* BCAction is “a national, grassroots advocacy and education organization” working to end breast cancer. It holds itself out as “the watchdog of the breast cancer movement.” *Id.* AARP is a “nonpartisan, nonprofit organization with a membership dedicated to addressing the needs and interests of people age fifty and older,” seeking to “enhance the quality of life for all by promoting independence, dignity, and purpose.” *Id.* at 6. AARP’s mission is focused, in part, on healthcare-related issues. *Id.*

B. Background on Genetics

Plaintiffs and Defendant generally do not dispute the core scientific principles underlying the genetics issues in this case. Here, the court relies upon expert declarations and testimony submitted by Plaintiffs and Defendant.

1. DNA

Genes are the units responsible for inheritance of discrete traits, such as the color of peas in a peapod. Kay Decl. at ¶ 15 (Dkt. 103); Tait Decl. at ¶ 32 (Dkt. 54). Genes are made from segments of deoxyribonucleic acid, or DNA. DNA is an integral component of chromosomes, the complex structures that carry genes and which are located within most cells of the human body. Pribnow Decl. at ¶ 18 (Dkt. 65); Kay Decl. from *AMP* Litigation at ¶ 131 (Dkt. 34-4). The human genome, the “whole of the genetic information of an organism,” is comprised of about 22,000 genes residing in 23 pairs of chromosomes. Tait Decl. at ¶ 32. Every cell in the human body contains a complete copy of the human’s genome.

DNA is a chemical compound containing within its molecular structure the genetic information necessary to code for most, if not all, aspects of embryogenesis, development, growth, metabolism, and reproduction. Pribnow Decl. at ¶ 22. At its most basic level, a DNA molecule is composed of five chemical elements: carbon, hydrogen, oxygen, nitrogen, and phosphorus. Kay Decl. at ¶ 12.

But DNA is unique from other molecules in that it encodes—provides the blueprint for—our highly organized, intricate, complex internal structures, and serves as the template for the complex molecules that allow us to extract, transform, and utilize the energy that is present in our environment. It can be said that DNA contains information necessary for all life functions. Aug. 23, 2013 Tutorial (Jackson) at 8:3-10, 8:22–9:18 (Dkt. 117); Nussbaum Decl. at ¶¶ 41-65

(Dkt. 61); Pribnow Decl. at ¶ 33; Pribnow 2nd Decl. at ¶¶ 21-24 (Dkt. 132). It is DNA's unique, informational aspect that sets it apart from other biological molecules. Tait Decl. at ¶ 32.⁶

The information in DNA is stored in the sequence of adjacent bases within the DNA strand through what is termed a "nucleotide sequence." Scientists often refer to DNA as a "polynucleotide," reflecting that DNA consists of a contiguous chain of chemical units called "deoxyribonucleotides." Pribnow Decl. at ¶¶ 22-23. The standard nucleotides in vertebrate DNA contain four different bases: adenine, thymine, cytosine, and guanine. As shorthand, scientists often denote nucleotides by the first letter of the names of their bases: "A" for adenine; "G" for guanine; "T" for thymine; and "C" for cytosine. These bases are linked together by chemical bonds via a sugar-phosphate backbone. Kay Decl. at ¶ 12. A DNA molecule is typically represented by the linear order of the nucleotide sequence.

Scientists can extract DNA from cells in the body. Such DNA is known as extracted "genomic" DNA or gDNA. Scientists can also chemically synthesize DNA. Whether genomic or synthetic, all DNA uses the same four nucleotides, and the information encoded in a specific nucleotide sequence is the same. Pribnow Decl. at ¶¶ 19-21, 27, 52-54; Pribnow 2nd Decl. at ¶¶ 5-11.

DNA often exists as a double helix, with two intertwined strands. This structure is made possible because each base in one strand is paired via hydrogen bonds with another base in the

⁶ In their Reply Brief, Plaintiffs dispute that DNA contains information, stating that "DNA and any segment of DNA do not contain 'information' (like computer or computer storage device or any analogous device)." Pls.' Reply Br. at 4 (Dkt. 98 at 16). But at the court's September 11, 2013 hearing, Myriad's corporate representative, Vice President for Technology Development and Laboratory Director, Dr. Ben Roa, testified that he agreed with the statements that "DNA can contain information," and that specifically "BRCA1 and BRCA2 genes do provide information relating to a patient's susceptibility to developing breast or ovarian cancer" Sept. 11, 2013 Hearing Tr. (Roa) at 105:9-17 (Dkt. 150).

other, complementary strand. Kay Decl. at ¶ 14. To better understand DNA's role as an informational molecule, one must understand the rules of base pair complementarity, or "Watson-Crick" base pairing, named after two of the scientists credited with deducing the structure of DNA upon recognizing the critical importance of base pair complementarity. Pribnow Decl. at ¶ 28. In Watson-Crick pairing, A pairs exclusively with T, while C pairs only with G. *Id.* at ¶¶ 28-29. The informational aspects of DNA are based on associations between the nucleotides that are governed by the natural law of Watson-Crick base pairing. Nussbaum Decl. at ¶¶ 41-65; Pls.' Reply Br. at 5 (Dkt. 98); Pribnow Decl. at ¶¶ 28-50.

A single-stranded DNA molecule has "directionality," meaning that the two ends of the molecule are chemically different. The "beginning" of a DNA molecule is called the 5' (5 prime)-end and the "end" of the molecule is called the 3' (3 prime)-end. Kay Decl. at ¶ 17. DNA also contains regions that can code for protein molecules. Protein-coding segments of native DNA are contained in "exons." *Id.* at ¶ 18. In humans, protein-coding exonic DNA sequences are typically interrupted by intervening DNA sequences known as "introns" that do not code for proteins, but may contain regulatory elements—which control when a cell activates a gene. *Id.*

DNA replicates through a complex process. Pribnow Decl. at ¶ 36. During replication, the DNA double helix is "unwound" and separated into single strands. *Id.* Single-stranded DNA binding proteins maintain DNA in its single-stranded conformation, preventing the strands from reassociating through Watson-Crick base pairing. Pribnow 2nd Decl. at ¶¶ 30-32. The two single strands of DNA then become templates for the synthesis of the strand that will form the opposite strand in a new double helix. Pribnow Decl. at ¶¶ 36-38; Pribnow 2nd Decl. at ¶¶ 13-17. DNA replication is "primed" by the presence of a short RNA primer that the enzyme responsible for

synthesizing new DNA – “DNA polymerase” – uses as a starting point to synthesize the new DNA strand. Pribnow 2nd Decl. at ¶¶ 13-17. The opposite strands are synthesized according to Watson-Crick base pairing rules, resulting in two identical copies of the original DNA sequence. *Id.* The entire genomic DNA of all human cells—all forty-six chromosomes’ worth (in twenty-three chromosome pairs)—is completely copied from end to end during each replication cycle. Pribnow Decl. at ¶ 37. Thus, all base pairs comprising the genome are exposed in an extended single-stranded form during each replication event. *Id.* In humans, this occurs trillions of times during the life of every person. *Id.* Both the integrity of the structure and the nucleotide sequence of each single strand of the entire double helix are critical for maintaining the fidelity of replication during the vast number of cell division events that occur. *Id.*

A change in the gene sequence is called a genetic “variant” or “polymorphism.” Any change, even to a single nucleotide, can constitute a variant. Some variants are harmless. Others, termed “mutations,” can cause disease or increase the risk of disease. Disease conditions in humans frequently are due to mutations in an individual’s copy of a single gene that gives rise to a protein different from the normal, or “wild-type,” protein expressed in persons without the disease. The genetic mutation that is responsible for such protein alteration may be determined and may be observed relatively easily through analysis of a person’s DNA sequence from a human sample. Tait Decl. at ¶ 34. These mutations can be found in exons or introns, although it is often easier for geneticists to identify disease-causing mutations in exons. Hence, mutation screens often concentrate on examining DNA sequences containing exons. Kay Decl. at ¶ 19.

In some instances, there is not enough information about a variant to classify it—the variant’s effect on the body is currently unknown. Such a variant is termed a “variant of unknown significance,” or “VUS.” The significance of the variant may be determined over time

though the collection and analysis of more data for that variant. Swisher Decl. at ¶ 40 (Dkt. 59); Nussbaum Decl. at ¶¶ 66-68. Through further investigation, most VUS results are ultimately reclassified as either deleterious or benign. The vast majority are reclassified as benign. Of course, the number of VUS reported is inversely proportional to the completeness of a genetic database. The more mutations discovered and characterized, the fewer VUS results will be returned. Nussbaum Decl. at ¶¶ 66-67.

2. RNA

RNA, or ribonucleic acid, is a chemical compound with four bases: guanine, cytosine, uracil, and adenine. Kay Decl. at ¶ 20. Thus, instead of DNA's thymine base, RNA contains uracil. *Id.* Common abbreviations of the RNA bases are: "G" for guanine, "C" for cytosine, "U" for uracil, and "A" for adenine. *Id.* Each base, together with one sugar and one phosphate molecule, makes up one repeating unit known as an RNA nucleotide. *Id.* Also like DNA, RNA is formed by a strand of bases that are linked together via a sugar-phosphate backbone. *Id.* The structures of the sugar-phosphate backbone of RNA and DNA, however, are different; while RNA contains a ribose sugar, the sugar component of DNA is a deoxyribose. *Id.* Because of these differences in structure, RNA usually exists as a single strand instead of the double helix associated with DNA. *Id.* DNA is generally more stable than RNA. *Id.*

RNA is generated in the body from DNA in a process called "transcription." *Id.* at ¶ 21. During transcription of RNA from DNA, a discrete segment of the DNA unwinds, and the bases of the DNA molecule act as "clamps" that hold the bases of the newly forming RNA in place while the chemical bonds of the sugar-phosphate backbone are formed. *Id.* This process is mediated by a structure in the cell known as RNA polymerase. *Id.*

A newly transcribed RNA molecule (transcript), or precursor messenger RNA (pre-mRNA), is processed to result in a mature messenger RNA (mRNA). *Id.* at ¶ 22. Pre-mRNA contains nucleotides that are eliminated during a process called “splicing.” *Id.* This involves splicing the introns out of the pre-mRNA, while the exons are ligated, or joined together, to form the intact mRNA molecule. *Id.*

3. Proteins

Proteins are generally large, complex molecules that play many critical roles in the body. *Id.* at ¶ 23. They are required for the structure, function, and regulation of the body’s tissues and organs. *Id.* Proteins are made up of hundreds or thousands of smaller units called amino acids, which are attached to one another in long chains. *Id.* There are 20 different amino acids that can be combined to make a protein. *Id.* The sequence of amino acids determines each protein’s unique 3-dimensional structure and its specific function. *Id.* Proteins are translated from mRNA through a process called “translation.” *Id.* at ¶ 24. During translation, mRNA serves as a template to assemble a protein. *Id.* Three consecutive bases in an mRNA molecule constitute a “codon,” which codes for one of the twenty amino acids. *Id.* Pairing interactions take place between an mRNA molecule and another RNA molecule known as tRNA, which serves as an adaptor during protein translation. *Id.* Specifically, sets of three nucleotides in the coding region of an mRNA react with three nucleotides in a tRNA in such a way as to cause the amino acid linked to the tRNA molecule to be chemically transferred to the growing polypeptide (a chain of amino acids linked together by peptide bonds) destined to become a protein. *Id.* The bases of the mRNA serve as “clamps” to hold the amino acids in place while the chemical bonds between the individual amino acids are formed. During translation, the mRNA template, the tRNA, the newly-forming polypeptide chain, and the next amino acid reside in a multi-protein complex

called a ribosome. *Id.* Once a protein is translated it typically undergoes post-translational or chemical modifications that are important for the protein's function. *Id.*

The genetic code describes which codons code for which amino acids. *Id.* at ¶ 25. For example, the codon adenine-thymine-guanine encodes the amino acid methionine. *Id.* Thus, the chemical composition of an mRNA molecule determines the amino acid composition of a protein. *Id.*

4. cDNA

Complementary DNA, or “cDNA,” is commonly synthesized from a mature mRNA in a reaction catalyzed by a protein known as reverse transcriptase. *Id.* at ¶ 26. cDNA is so named because each base in the cDNA can bind to a base in the mRNA from which the cDNA is synthesized. *Id.* In other words, it is “complementary” to the mRNA from which it is synthesized. *Id.* cDNA can be structurally different from native DNA. Most notably, cDNA made from an mRNA does not contain introns. *Id.* at ¶ 27. DNA generally contains intronic sequences—although DNA fragments may contain only exons. cDNA is also functionally different from DNA. *Id.* at ¶ 28. Most critically, DNA contains regulatory sequences. *Id.* These regulatory sequences are not present in cDNA because they are not present in the mRNA from which the cDNA is synthesized. *Id.*

5. Primers and Probes

A primer is a short, synthetic, single-stranded DNA molecule that binds specifically to an intended target nucleotide sequence. *Id.* at ¶ 29. The sequence of the primer is necessarily complementary to the target sequence, so that the bases of the primer and the bases of the target sequence bind to each other. *Id.* In human genetic testing, primers bind to human gene sequences that are an exact match according to the law of Watson-Crick base pairing. Pribnow

Decl. at ¶ 91. Binding a primer to its target sequence is the first step in amplifying a segment of DNA—the production of multiple copies of a specific DNA segment for DNA sequencing reactions or other molecular characterization. *Id.*; Kay Decl. at ¶¶ 29-30.

Scientists create primers. In so doing, they consider primer size and other aspects, such as the exact portion of the DNA segment targeted. Pribnow 2nd Decl. at ¶ 8. These considerations are dictated by the nucleotide sequence of the DNA segment to which the primer is intended to bind. *Id.* For example, if the targeted sequence is a naturally occurring BRCA1 or BRCA2 sequence, the starting points for primer creation necessarily must be the complement of the naturally occurring BRCA1 or BRCA2 sequences flanking the specific DNA region the scientist wishes to amplify. *Id.* Because the primers in a pair are designed to “hybridize” to their BRCA primer binding sites per Watson-Crick base pairing rules, the BRCA primers must contain sequences identical to the BRCA sequence directly opposite its binding sites. *Id.*

Typical primer pairs used in the most common method of DNA amplification, polymerase chain reaction (*see* discussion of polymerase chain reaction *infra* Part I.B.6.), are between 15 to 18 nucleotides or 25 to 30 nucleotides in length. Roa Decl. at ¶ 16 (Dkt. 63); Pribnow Decl. at ¶ 91. To hybridize well to a person’s DNA, the length of the nucleotide sequence of the primer that binds to the sample DNA should be at least 15 nucleotides long. Roa Decl. at ¶¶ 16, 21-22.

In addition to sequences that are identical to naturally occurring DNA sequences, primers may have additional appended sequences on their ends, such as “Next Generation Sequencing (NGS) adaptor sequences.” These adaptor sequences do not hybridize to the targeted genetic sequence. Elliott Decl. ¶¶ 15-17 (Dkt. 47); Elliott 2nd Decl. at ¶¶ 4-5 (Dkt.

136). Neither do they affect a primer pair's function in hybridizing to portions of a targeted DNA sequence. The primer pairs bind to and prime the same portion of the DNA sequence regardless of the presence or makeup of any such appended molecule. Elliott Decl. at ¶¶ 15-17; Elliott 2nd Decl. at ¶¶ 4-5.

Genetic testing methods can also utilize "probes." Pribnow Decl. at ¶ 85. Probes are similar to primers in that they are short segments of DNA that are capable of hybridizing to a DNA segment according to the rules of Watson-Crick base pairing. *Id.* A probe is used to detect the presence or absence of a particular DNA sequence in a DNA sample. Tait Decl. at ¶ 26. Thus, as with primers, the composition of a probe is dictated by the DNA sequence a scientist wants to identify. *Id.* at ¶ 24. The probe's DNA sequence is a complement to the sequence of DNA the probe will be used to detect, so that the probe will hybridize to the DNA target through Watson-Crick base pairing. Pribnow Decl. at ¶¶ 85-87; Tait Decl. at ¶¶ 22-26.

Primers and probes may utilize sequences that can hybridize to the sequence that would be present in a cDNA of the gene. In other words, primers and probes can hybridize to exonic-only sequences. But, primers and probes are not cDNA. Pribnow Decl. at ¶ 86. cDNA is typically not used as a primer or probe. *Id.* at ¶ 84. In the genetic testing that Plaintiffs contend their patent claims cover, cDNA may not be used much at all. *Id.* Rather, one simply amplifies a segment of DNA and uses it to interrogate a gene for medically important mutations. *Id.* cDNA is not typically used as a probe or primer in genetic testing in part because it is too large. *Id.* In addition, primers often are designed to hybridize to noncoding regions of the gene (introns) in order to copy the sequence of the intron immediately adjacent to the exon, in addition to the exon itself, thus mirroring the native nucleotide sequence. *Id.* Since cDNA does not contain introns, it cannot be used as a primer in this application. *Id.* Genomic DNA extracted

from the body is not typically used as primers or probes, although it would be possible to do so.

Id. at ¶ 88.

Using probes and primers in genetic testing, including BRCA testing, does not fundamentally change the DNA that is analyzed. Pribnow Decl. at ¶ 87. More specifically, primers and probes do not alter the underlying, naturally occurring DNA sequence that is being read. *Id.* Therefore, they do not alter the underlying DNA's functional properties or identity for the purposes of genetic testing. *Id.*

At the time of Myriad's patents, the techniques for creating primers were well known in the art. Primers were generally created using commercially available "oligonucleotide synthesizing machines." *See* '282 Patent col.16 ll.43-48 (BRCA1); '492 Patent col.15 ll.30-37 (BRCA2).

As with primers, the creation and use of DNA probes in genetic testing experiments was well known and widely used prior to August 1994, when Myriad submitted its application for the first of the patents at issue in Plaintiffs' Motion. Tait Decl. at ¶ 26; '282 Patent col.15 ll.9-20, col.17 ll.15-32, col.21 ll.34—col.22 l.25. Probe hybridization results from Watson-Crick base pairing between two complementary strands of nucleic acids. Hybridization, a form of binding between molecules, occurs as a result of the inherent chemical properties of nucleic acid molecules and gives double-stranded DNA its characteristic helical structure. Tait Decl. at ¶ 22.

6. Polymerase Chain Reaction

Many copies of an input DNA are required to sequence genes. Those who want to sequence and test human genes utilize methods for amplifying—creating copies—of a segment of genomic DNA products. Kay Decl. at ¶ 31; Pribnow Decl. at ¶ 17. Whether produced in a laboratory or by nature, amplified DNA is indistinguishable from the original DNA that was

copied, both in its chemical structure and, importantly, the sequence information contained in the DNA. Pribnow Decl. at ¶ 19.

The most widely used DNA amplification method is the polymerase chain reaction (PCR). Kay Decl. at ¶ 32. PCR mimics the processes of DNA replication in the cell. Pribnow Decl. at ¶¶ 55-59; Pribnow 2nd Decl. at ¶¶ 13-17. When PCR is used in conjunction with a targeted segment of genomic DNA, numerous exact duplicates are synthesized, and these are indistinguishable in sequence and chemical composition from the targeted genomic DNA. *Id.*

PCR was developed in the 1980s by Dr. Kary Mullis at Cetus Corporation to develop exact duplicates—“amplicons”—of DNA segments. Pribnow Decl. at ¶ 16; Tait Decl. at ¶¶ 29-31. In 1989, the publication *Science* identified PCR and its use of a DNA polymerase from a thermophilic bacterium, *Thermus aquaticus* (*Taq* DNA polymerase), as the “Molecule of the Year,” and Dr. Mullis won the Nobel Prize in Chemistry for his invention in 1993. Tait Decl. at ¶ 29. Thus, as the asserted patents acknowledge, PCR was a well-understood and routine activity in the scientific community prior to the time Myriad filed its August 1994 application corresponding to those asserted patents, and prior to the identification of the BRCA1 or BRCA2 gene sequences. *Id.* at ¶ 31; ’441 Patent col.17 ll.21-37.

In any presently known process for analyzing a human’s genes, the first step is to obtain a person’s blood, saliva, or a cultured cell sample. DNA is then extracted from this sample. This extracted, genomic DNA represents a person’s “diploid” genome, as it contains two copies of each autosomal (non sex chromosome) gene. Roa Decl. at ¶ 3; Pribnow Decl. at ¶ 55. The genomic DNA is then “fragmented,” or cut into small pieces, often through sonication or biochemical shearing. The fragmentation randomly cuts all parts of the DNA into many

randomly sized pieces. These fragments are typically about 1000 nucleotides long, but smaller fragments can be created. Roa Decl. at ¶ 4; Pribnow Decl. at ¶ 55.

The PCR process begins by mixing the fragmented genomic DNA with: 1) a thermostable DNA polymerase enzyme; 2) a pool of all four DNA nucleotides (A, C, T and G); and 3) a great excess of single-stranded primer pairs. Kay Decl. at ¶ 32; Pribnow Decl. at ¶¶ 55-59; Pribnow 2nd Decl. at ¶¶ 13-17; Tait Decl. at ¶¶ 23-29. One primer in the pair is complementary to one end of the region to be amplified on one strand of the template DNA molecule and the other primer in the pair is complementary to the other end of the region to be amplified on the other strand of the template DNA. Kay Decl. at ¶ 32.⁷

Next, several steps occur in a cyclical reaction, as depicted in the illustration below. First, the template-primers mixture is heated so that the bonds linking the two strands of the template DNA molecule are overcome, causing the strands to separate. *Id.* at ¶ 33. This is called denaturation. *Id.* Second, the mixture is cooled enough to allow one copy of each primer to bind to its complementary template DNA sequence, in a process called annealing. *Id.* Third, the DNA polymerase adds nucleotides to the 3'-end of each of the primers in an order complementary to the template DNA. *Id.* This extension reaction results in the generation of a

⁷ Some of the exon-only sequences in the BRCA1 and BRCA2 genes are so long that, when the entire gene is fragmented for PCR, some of the resulting fragments consist only of exons (i.e., they have no intron fragments). For instance, Exon 11 of BRCA1 is about 3,400 nucleotides long, while an exon segment in BRCA2 is about 5,000 nucleotides long. Roa Decl. at ¶ 14. Thus, the nucleotide sequence in the primer pairs necessary to amplify those exon fragments, and the resulting amplified DNA molecules or “amplicons,” will inevitably share sequence similarity only with part of an exon in the gene. *Id.* at ¶¶ 13-15. Such large exons are likewise “too large to be encompassed within a single amplicon due to inherent limitations in the PCR process.” As a result, the primer pairs are designed “to yield multiple distinct amplicons across the length of the gene” and, for large exons, multiple amplicons are tiled across the exon. *Id.* at ¶¶ 13-15.

The diagram illustrates the three steps of PCR: ① Denaturation, ② Annealing, and ③ Elongation. It shows how a single DNA molecule is amplified exponentially through multiple cycles. In the first cycle, the double-stranded DNA is heated to separate into two single strands. In the second cycle, primers (green circles) bind to the single strands, and DNA polymerase (green lines) extends them to create two new double-stranded molecules. This process repeats, leading to exponential growth of the DNA fragments.

19

patient's sequence of nucleotides, as reflected in the amplicons generated from her DNA. Aug. 23, 2013 Tutorial Tr. (Jackson) at 13:4-11, 23:16 – 24:3, 30:6 – 31:13; Sept. 11, 2013 Hearing Tr. (Roa) at 112:4-17; Pribnow 2nd Decl. at ¶¶ 33-35. For example, whether the information in a woman's BRCA1 or BRCA2 genes predisposes her to an increased risk of hereditary breast or ovarian cancer can be determined by analyzing the sequence of at least portions of her BRCA1 and BRCA2 genes. The amplicons generated during PCR enable this evaluation. Sept. 11, 2013 Hearing Tr. (Roa) at 105:1-15, 111:20-25, 112:1-15; Pribnow Decl. at ¶¶ 64-70; Pribnow 2nd Decl. at ¶¶ 33-35.

7. Sequencing

After PCR, the resulting amplified DNA can be sequenced. This means that the specific nucleotide (adenine (“A”), thymine (“T”), cytosine (“C”), or guanine (“G”)) in each position of the DNA is identified or “read.” The identified sequence of an individual person's gene or genes is commonly called a “germline” sequence, meaning the gene sequence that a person inherited at birth. Roa Decl. at ¶ 24.

Two types of sequencing are at issue in this case: dideoxy sequencing (also known as Sanger sequencing) and Next-Generation Sequencing (NGS). Both types mimic DNA cell replication by using primers, DNA polymerase, and nucleotides—some of which have been chemically modified, but in ways that do not alter their Watson-Crick pairing functions. Aug. 23, 2013 Tutorial Tr. (Jackson) at 24:4-25; Elliott Decl. at ¶¶ 23-31; Tait Decl. at ¶¶ 35-36.

Sanger sequencing was developed in 1977 and is named for its inventor, Frederick Sanger.⁸ NGS was not developed until the 2000s. Aug. 23, 2013 Tutorial Tr. (Roa) at 26:20 – 27:1; Sept. 11, 2013 Hearing Tr. (Roa) at 116:4-8; Tait Decl. at ¶¶ 35-37. Defendant’s testing employs both Sanger sequencing and NGS. Elliott Decl. at ¶¶ 23-31; Elliott 2nd Decl. at ¶¶ 14-16; Tait Decl. at ¶¶ 35-36.

By the time Myriad submitted its first application corresponding to the asserted patents in August 1994, the laboratory techniques used to accomplish hybridization, amplification, and sequencing for the purpose of observing a genomic, or “native” gene sequence in a human sample were well understood, widely used, and fairly uniform insofar as any scientist engaged in obtaining the sequence of a gene in a human sample would likely have relied on the same techniques and general approach. Tait Decl. at ¶ 37. Likewise, the laboratory materials, reagents, and protocols to accomplish these tasks were well known and widely available in the art by that time, as the asserted patents acknowledge. Tait Decl. at ¶ 31; ’441 Patent col.17 ll.20-27 (“These methods are well known and widely practiced in the art.”). Plaintiffs’ patents state that “the practice of the present invention employs, unless otherwise indicated, conventional techniques of chemistry, molecular biology, microbiology, recombinant DNA, genetics, and immunology.” *See, e.g.*, ’282 Patent col.25 ll.50-55.

a. Sanger Sequencing

Sanger sequencing includes, among other things, the application of artificial DNA nucleotide analogues known as dideoxynucleotide chain terminators to the amplified DNA.

⁸ Mr. Sanger passed away on November 19, 2013. The sequencing method bearing his name was used to sequence the first human genome, the first draft of which was completed in 2000 after thirteen years of work. Sept. 11, 2013 Hearing Tr. (Roa) at 115:19-25; Elliott Decl. at ¶ 21.

Either the primers or the terminators have labels or tags, such as fluorescent dyes or radioactive phosphorus, that can be read during sequencing of the DNA to identify what type of nucleotide is present at each position in the gene. Roa Decl. at ¶ 25; Elliott Decl. at ¶¶ 35-39. Improvements in the Sanger method by 1987 led to automation of the sequencing process. These Sanger sequencing methods were widely used well before the first of Myriad's August 1994 applications corresponding to the asserted patents had been filed, and prior to the identification of the BRCA1 or BRCA2 gene sequences. Tait Decl. at ¶ 36.

b. Next Generation Sequencing (NGS)

NGS does not rely on the termination of DNA synthesis for resolution of a DNA's sequence. Unlike Sanger sequencing, NGS does not use dideoxy sequencing. Elliott Decl. at ¶¶ 24-25. Instead, DNA molecules are "sequenced by synthesis" through a process where tagged (e.g., fluorescent) nucleotides are added to a growing synthetic DNA molecule in a controlled, stepwise fashion. As nucleotides are incorporated into the new synthetic molecules, the corresponding tag is released. This release is detected by a machine utilizing a technology similar to a camera that detects the flash of the fluorescent tag. Each nucleotide (A, T, C, and G) has a unique color fluorescent tag attached so that the machine can detect which nucleotides were added, and in which order. Roa Decl. at ¶ 26; Elliott Decl. at ¶¶ 20, 24-25, 29-31.

8. Large Rearrangement Analysis

In addition to gene sequencing techniques like Sanger sequencing and NGS, scientists may also employ "large rearrangement analysis." This analysis enables scientists to determine if a person's gene contains large nucleotide deletions or sequence duplications which may not be observed during the sequencing processes described above.

Two common examples of large rearrangement analysis are multiplex ligation-dependent probe amplification (MLPA) and microarray. Roa Decl. at ¶ 30; Elliott Decl. at ¶¶ 40-42. Both MLPA and microarray processes use probes with a series of nucleotides precisely complementary to a piece of a single strand of the gene to be analyzed, such as BRCA1 or BRCA2 genes. MLPA uses pairs of probes that hybridize to adjacent segments of the target gene sequence and, after being hybridized, are chemically fused or “ligated” to form a single molecule. MLPA probes are chosen to be complementary to an “allele” in a gene, which refers to a form of the gene having a certain specific nucleotide sequence, such as a wild-type sequence, variant sequence, or mutation sequence of interest, and thus may be called a “wild-type” or a “mutated” allele. For large rearrangement testing, probes are targeted to parts of the gene that may contain deletions or duplications and are typically designed to detect deletion or duplication of one or more exons. Roa Decl. at ¶ 31; Elliott Decl. at ¶¶ 42-44, 47-52.

In the MLPA process, multiple synthetic probes are used containing various specific nucleotide sequences that target sequences of the parts of the gene to be analyzed (e.g., typically one pair of probes for each exon). The probes also include generic “primer tail” nucleotide sequences that allow for PCR probe amplification. MLPA probes are designed to assess large deletion or duplication mutations in or near coding exons in the gene. MLPA uses pairs of probes containing target gene sequences that are adjacent to each other. The probes hybridize to any of the DNA fragments that are complementary to those probes. Matching probes that hybridize next to each other are then ligated to form a longer oligonucleotide. Because the probes also have primer tail sequences, the resulting ligated DNA is then amplified through PCR and labeled with fluorescent tags. Roa Decl. at ¶ 32; *see* Elliott Decl. at ¶¶ 43-46. The resulting amplified MLPA products are analyzed and compared using computer software.

By comparing the relative copy number of MLPA products in a patient against a wild-type control, the presence of a large deletion or duplication can be detected. For example, if certain probes hybridize at approximately 50 percent the amount the hybridization obtained in a wild-type control, it means that there was no section of the gene to which those probes were complementary in one of the patient's expected two copies of the gene, indicating a deletion in the relevant region in one of the patient's copies of the gene. Conversely, a 50 percent increase in MLPA probes in a certain region indicates duplication of that region in the gene. Roa Decl. at ¶ 33; *see* Elliott Decl. at ¶¶ 46-48.

The hybridization of the probes is detected and quantified by amplification of the ligated longer probe. Conversely, if the region where the probe would normally hybridize has been deleted on one or more of the patient's chromosomes, then less-than-expected hybridization and ligation will take place and less than the expected amount of amplification will result. In this way, detecting hybridization, or the lack of hybridization, allows scientists to compare a patient's DNA sequence to a wild-type sequence and determine whether mutations are present. For this reason, using MLPA to detect an alteration in DNA necessarily requires detection of the wild-type allele through hybridization of the probes. Roa Decl. at ¶ 34; *see* Elliott Decl. at ¶¶ 46-48.

"Microarray," another form of large rearrangement analysis, uses a solid surface, such as a glass slide, with a collection of microscopic spots to which different DNA probes are attached. A microarray process employing comparative genomic hybridization (microarray-CGH) uses patient genomic DNA that is fragmented into small pieces. Synthetic products can be generated from the fragmented patient DNA by primer extension and labeling with a specific fluorescent dye. Similar products can be generated from fragmented wild-type DNA that is labeled with a different fluorescent dye. Alternatively, genomic fragments can be directly labeled with

fluorescent dyes. A mixture containing equal amounts of differentially labeled products representing patient genomic and wild-type DNA are hybridized to the microarray slide with immobilized probes tiled across the entire coding region of BRCA1 and BRCA2. Roa Decl. at ¶ 35; *see* Elliott Decl. at ¶¶ 51-53.

Following hybridization, the microarray slides are scanned and the relative dye intensities are analyzed and quantified. Equal amounts of the two dye signals indicate a normal result or, in other words, show that no large rearrangement has been detected. In contrast, if there is a relative decrease in the amount of patient's dye signal relative to wild-type dye signal, then deletion in one copy of the BRCA1 or BRCA2 gene regions covered by the affected probes is indicated. Conversely, if there is a relative increase in a patient's dye signal, then that result indicates a duplication of the gene region corresponding to those probes. The resulting data obtained in the microarray-CGH analysis allows for identification of large genomic deletions or duplications in the BRCA1 or BRCA2 genes. These large rearrangements can occur anywhere in the gene, and may involve a single exon, multiple exons, or even the entire gene coding region. Roa Decl. at ¶ 36; *see* Elliott Decl. at ¶ 53.

Microarray-CGH necessarily requires hybridization of sample DNA to a probe specific for the gene of interest, such as BRCA1 or BRCA2, and detection of that hybridization product. The resulting data allows identification of both the existence of the allele of interest and the existence of large mutations in a patient's germline sequence, i.e., comparing the sequence of the patient's BRCA1 or BRCA2 gene to wild-type, by comparing probe hybridization relative to the wild-type BRCA1 or BRCA2. Because the microarray process is performed by comparing the hybridization of the patient's allele to the hybridization of the wild-type allele,

this process necessarily requires detection of the wild-type allele through hybridization. Roa Decl. at ¶ 37.

C. The Race to Locate and Sequence BRCA1 and BRCA2

Breast cancer is by far the most often diagnosed type of cancer among women, affecting about one in eight women. Swisher Decl. at ¶ 19. Among the entire population of men and women combined, breast cancer is the second most diagnosed cancer.⁹ The National Cancer Institute (NCI) estimates approximately 232,340 new cases of female breast cancer and 2,240 new cases of male breast cancer will have been diagnosed in the United States in 2013. *Id.* The NCI estimates that breast cancer will have caused approximately 39,620 female deaths and 410 male deaths in the United States in 2013. *Id.* Ovarian cancer is the eighth most common cancer in women. Although less common than breast cancer, it causes more deaths in the Western world than any other gynecologic cancer. *Id.* at ¶ 21.

In the 1980s, breast cancer patients mobilized to increase public awareness of the breast cancer epidemic. Due to these efforts and those of breast cancer organizations, the Department of Defense created a research program devoted to breast cancer research. Between 1990 and 2008, the annual funding for this research increased from \$90 million to \$2.1 billion. Parthasarathy Decl. from *AMP* Litigation at ¶ 10 (Dkt. 34-7).

Also during the 1980s, scientists from the United States, England, France, Germany, Japan, and other countries were competing to first identify the nucleotide sequences linked to breast cancer. In 1989, various European and American research laboratories participated in an International Breast Cancer Linkage Consortium. *Id.* at ¶ 11.

⁹ See <<http://www.cancerresearchuk.org/cancer-info/cancerstats/world/breast-cancer-world/>>.

In 1990, a research group led by Mary-Claire King at the University of California, Berkeley announced that it had discovered that the Breast Cancer Susceptibility Gene 1 (BRCA1) was located on chromosome 17. With this discovery, research teams around the world intensified efforts to be the first to sequence the BRCA1 gene. *Id.* at ¶ 11. Among them were different teams led by Dr. King; Dr. Mark Skolnick, co-founder of Myriad; and Dr. Michael Stratton of the Institute for Cancer Research, London (ICR). *Id.*

In September 1994, Dr. Skolnick's group at Myriad—including researchers from the National Institute for Environmental Health (NIEH), an agency of the National Institutes of Health (NIH)—announced that they had sequenced the BRCA1 gene. *Id.* at ¶ 11. They won a hard-fought “race,” as journalists reported at the time:

The race to find the breast-cancer gene has been one of the most closely-watched and publicized of a host of gene hunts in recent years. The pursuit of the gene was triggered in late 1990 when Mary-Clare King, a geneticist at the University of California at Berkeley, stunned the cancer-research community by pinpointing the the [sic] gene's approximate location. About a dozen laboratories around the world, including Dr. King's, have been intensely probing a tiny region of genetic material since then. In the past few months, scientists said they had identified about 30 genes in the approximate region but had pared the search down to about four to six likely culprits. Dr. Skolnick said the first hint they had latched onto the gene came about two months ago. Since then they have worked to identify its structure.

Scientists Say They've Found Gene That Causes Breast Cancer, The Wall Street Journal, September 14, 1994 (Dkt. 114-2). By the time this discovery was publicly announced in September 1994, Myriad had applied for patents related to BRCA1, including the '282 and '441 Patents.¹⁰

After the sequencing of BRCA1, many scientists thought there was at least one more gene linked with breast cancer, and the search for that gene continued. *Id.* at ¶ 12. By 1994, the

¹⁰ These patents have the priority date of August 12, 1994, and begin to expire in August 2014.

existence of a 'BRCA2 gene' and its location on chromosome 13q was known, as described by

Plaintiffs in the background of the invention section of the '441 Patent:

Intense efforts to isolate the BRCA1 gene have proceeded since it was first mapped in 1990. [citations omitted]. A second locus, BRCA2, has recently been mapped to chromosome 13q (Wooster et. al., 1994) and appears to account for a proportion of early-onset breast cancer roughly equal to BRCA1, but confers a lower risk of ovarian cancer.

'441 Patent col.2 l.46—col.3 l.4.

In December 1995, the group led by Dr. Mark Stratton announced they had mapped and sequenced the elusive second gene, the BRCA2 gene, which was linked to incidence of ovarian cancer, as well as female and male breast cancer. Parthasarathy Decl. from *AMP* Litigation at ¶ 12. The day before the Stratton group published the BRCA2 gene sequence in the scientific journal, *Nature*, however, Myriad announced that it too had found the BRCA2 gene. *Id.* Myriad submitted its sequence to GenBank, an international depository of gene sequence information, and applied for patents on the BRCA2 gene in the United States and Europe. *Id.*

The sequencing of the BRCA1 and BRCA2 genes were landmark events in genetics, as mutations in these genes are responsible for many breast and ovarian cancer cases. About 10 percent of breast cancers are inherited genetically, about 5 percent as a result of a BRCA1 or BRCA2 genetic mutation. Swisher Decl. at ¶ 20. Individuals with BRCA1 and BRCA2 mutations have about a 45 to 87 percent risk of developing breast cancer by age 70. *Id.* Between 20 to 25 percent of ovarian cancers are inherited genetically. For women with inherited ovarian cancer, 75 percent of them can attribute the cancer to either BRCA1 or BRCA2. About 50 percent of inherited ovarian cancers are caused by BRCA1 mutations, about 25 percent are caused by BRCA2 mutations, and the remaining 25 percent are caused by other genes. Women with inherited BRCA1 mutations have a 40 to 52 percent cumulative risk of ovarian cancer by

the time they reach 70 years old. For women with inherited BRCA2 mutations, the risk is approximately 15 to 25 percent. *Id.* at ¶ 22. Very little can be done for patients once diagnosed with ovarian cancer, making preventive care vital. *Id.* at ¶ 23.

D. Myriad's Testing Products: BRACAnalysis, BART, and myRisk

Beginning in 1994 and continuing for several years, Myriad obtained numerous patents related to BRCA1 and BRCA2. By 1996, it began to market BRCA1 and BRCA2 molecular testing products. That year, Myriad introduced BRACAnalysis, a molecular diagnostic test used to detect the presence and characterization of 'point' or small mutations in the BRCA1 or BRCA2 gene that are responsible for a majority of hereditary breast and ovarian cancers. Ford Decl. at ¶¶ 1, 3; Sept. 12, 2013 Hearing Tr. (Ford) at 312:5-25 (Dkt. 151). The BRACAnalysis test does not include large rearrangement testing for BRCA1 and BRCA2—testing that can identify initially false negative results in a BRACAnalysis point mutation test.

Myriad offers an additional test that provides large rearrangement testing for the BRCA1 and BRCA2 genes, called BRACAnalysis Rearrangement Test, or "BART." Sept. 12, 2013 Hearing Tr. (Ford) at 312:19-25. But a patient who obtains Myriad BRACAnalysis testing does not automatically get follow-up BART testing. *Id.* at 101:22-52:15; Swisher Decl. at ¶¶ 97-98; Matloff Decl. at ¶ 7 (Dkt. 49). If a patient does not satisfy Myriad's criteria for being at high risk of a large rearrangement mutation in her BRCA1 and BRCA2 genes, or if her insurance does not cover BART, then the patient must pay for the BART test separately. Sept. 12, 2013 Hearing Tr. (Ford) at 314:2-9. According to a 2013 peer-reviewed study in the *Journal of Clinical Oncology*, Myriad's criteria for providing BART large rearrangement testing automatically as part of its BRCA testing does not cover half the patients who have large rearrangement mutations: "[f]ewer than half of the large rearrangement carriers in the present study met Myriad Genetics

Laboratories' criteria . . . for automatic large rearrangement testing." Chao Decl. at Exh. P at 212 (Weitzel *et al.*, 31(2) J. Clin. Oncol. 210-06 (2013)).

The current list price for BRACAnalysis is \$3,340 and the list price for BART is \$700. Sept. 12, 2013 Hearing Tr. (Ford) at 313:11-15. Together, the list price for both tests is \$4,040. Ford. Decl. at ¶ 11. Those who get BART in addition to BRACAnalysis are billed separately for the two tests. Pls.' Reply Br. at 135. But not all of Myriad's patients have insurance coverage for BART. Mark C. Capone, President of Myriad Genetic Laboratories, Inc., suggested on May 7, 2013, that approximately 20 percent of patients receiving BRACAnalysis did not have insurance coverage for BART. Sept. 12, 2013 Hearing Exh. 3 at 15 (Dkt. 144-2) ("Our Managed Care team continued to make significant progress on BART reimbursement in the fiscal third quarter and we ended the quarter with reimbursement coverage for approximately 80 percent of patients."). Because BART is not automatically included as part of Myriad's BRCA1 and BRCA2 testing, some patients must pay the \$700 out-of-pocket for BART if they are to get that testing.

On September 5, 2013, Myriad announced a limited launch of myRisk, a new "multi-gene diagnostic test that will provide increased sensitivity by analyzing 25 genes associated with eight major cancers including: breast, colorectal, ovarian, endometrial, pancreatic, prostate, gastric and melanoma." Myriad Press Release, Sept. 5, 2013.¹¹ This test utilizes the Next-Generation Sequencing used by Defendant and tests for 24 of 25 of the same genes as Defendant's CancerNext panel. Chao Decl. at ¶¶ 46-47 and Exh. I. In an investor and analyst presentation given on May 9, 2013, Myriad described myRisk as a "significant improvement of BRACAnalysis." *Id.* at ¶ 45, Exh. I at 19-45. Similarly, in a press release dated May 30, 2013,

¹¹ See <<http://investor.myriad.com/releasedetail.cfm?ReleaseID=788983>>.

Myriad stated: “myRisk represents a scientific advancement that will revolutionize hereditary cancer testing for appropriate patients.” *Id.* at Exh. J at 1.

For the time being, it appears that this launch provides limited access to the public, as it is for “a limited number of medical and scientific thought leaders”:

myRisk Hereditary Cancer is being launched in a phased approach beginning with an early-access, clinical-experience program to a limited number of medical and scientific thought leaders followed by an expanded access program later in the year. The Company will present extensive clinical validity data for myRisk Hereditary Cancer at The Collaborative Group of the Americas on Inherited Colorectal Cancer (CGA) annual meeting in October and the San Antonio Breast Cancer Symposium in December.

Myriad Press Release, Sept. 5, 2013.

E. The AMP Litigation

1. Judge Sweet’s Decision—Southern District of New York

A group of medical organizations and individuals sued Myriad in 2009, challenging fifteen composition and method claims in seven of Myriad’s BRCA1 and BRCA2-related patents on the grounds that they were drawn to products of nature and mental processes—subjects that are patent ineligible under 35 U.S.C. § 101. *Association for Molecular Pathology v. United States Patent and Trademark Office*, 702 F. Supp. 2d 181, 186 (S.D.N.Y. 2010).¹² These patents were drawn to “(1) isolated DNA containing all or portions of the BRCA1 and BRCA2 gene sequence and (2) methods for ‘comparing’ or ‘analyzing’ BRCA1 and BRCA2 gene sequences to identify the presence of mutations correlating to breast or ovarian cancer.” *Id.* at 184. United States District Court Judge Robert W. Sweet characterized the overarching issue

¹² Specifically at issue in *AMP* were claims 1, 2, 5, 6, 7, and 20 of the ’282 Patent; claims 1, 6, and 7 of the ’492 Patent; claim 1 of U.S. Patent 5,693,473 (the ’473 Patent); claim 1 of the ’999 Patent; claim 1 of U.S. Patent 5,710,001 (the ’001 Patent); claim 1 of the ’441 Patent; and claims 1 and 2 of the ’857 Patent.

presented as: “[a]re isolated human genes and the comparison of their sequences patentable?”

Id. at 185. Judge Sweet answered that question in the negative and granted summary judgment in favor of the plaintiffs. *Id.* at 185.

Judge Sweet construed “isolated DNA” to mean a “segment of DNA nucleotides existing separate from other cellular components normally associated with native DNA, including proteins and other DNA sequences comprising the remainder of the genome, and includes both DNA originating from a cell as well as DNA synthesized through chemical or heterologous biological means.” *Id.* at 217. He concluded that neither Myriad’s isolated DNA (composition) claims, nor its method claims were drawn to patent eligible subject matter.

The composition claims turned on the issue of “whether or not claims directed to isolated DNA containing naturally-occurring sequences fall within the products of nature exception to § 101.” *Id.* at 220. As a starting point, Judge Sweet noted that “Supreme Court precedent has established that products of nature do not constitute patentable subject matter absent a change that results in the creation of a fundamentally new product.” *Id.* at 222. Even “‘purification’ of a natural compound, without more, is insufficient to render a product of nature patentable.” *Id.* at 223. Judge Sweet rejected Myriad’s argument that “purified DNA” is necessarily patent eligible because DNA doesn’t exist in nature in a purified form. *Id.* at 224. Observing that even a “purified product” must have “‘markedly different characteristics’ in order to satisfy the requirements of § 101,” Judge Sweet analyzed whether Myriad’s “isolated DNA” had “‘markedly different characteristics’ from a product of nature.” *Id.* at 227-28 (quoting *Diamond v. Chakrabarty*, 447 U.S. 303, 310 (1980)).

Myriad, citing the “chemical nature of DNA,” argued that isolated DNA “is ‘markedly different’ from DNA found in nature” due to “structural and functional” differences. *Id.* at 228.

For example, Myriad pointed out that there are chromosomal proteins associated with native DNA that are not associated with isolated DNA. *Id.* at 229-30. But Judge Sweet found that focusing on these differences ignores the reason DNA is unique from other chemical compounds in nature—because it encodes and conveys information:

The information encoded in DNA is not information about its own molecular structure incidental to its biological function, as is the case with adrenaline or other chemicals Rather, the information encoded by DNA reflects its primary biological function: directing the synthesis of other molecules in the body—namely, proteins. . . . DNA, and in particular the ordering of its nucleotides, therefore serves as the physical embodiment of laws of nature—those that define the construction of the human body. . . . Consequently, the use of simple analogies comparing DNA with chemical compounds previously the subject of patents cannot replace consideration of the distinctive characteristics of DNA.

Id. at 228-29. It is DNA’s nucleotide sequence that is critical to both its “natural biological function” and “the utility associated with DNA in its isolated form.” *Id.* at 229.

Judge Sweet also rejected Myriad’s argument that isolated DNA is distinct from native DNA insofar as it “may be used in applications for which native DNA is unsuitable, namely, in ‘molecular diagnostic tests (e.g., as probes, primers, templates for sequencing reactions), in biotechnological processes (e.g. production of pure BRCA1 and BRCA2 protein), and even in medical treatments (e.g. gene therapy).’” 702 F. Supp. 2d at 230-31 (quoting Myriad’s *AMP* Reply Br. in Support of Motion for Summary Judgment at 9 (other citations omitted)). Judge Sweet noted that the cited applications depend on the single-stranded isolated DNA segment having “the identical sequence as the complementary DNA strand to the DNA strand containing the target DNA sequence.” *Id.* at 231, n.54. The BRCA-specific nucleotide sequence is “the defining characteristic of the isolated DNA that will always be required to provide the sequence-specific targeting and protein coding ability that allows isolated DNA to be used for the various applications cited by Myriad.” *Id.* at 232.

Judge Sweet found unpersuasive Myriad's contention that it "created" the claimed BRCA DNA molecules when it identified the "specific segments of chromosomes 17 and 13 that correlated with breast and ovarian cancer (BRCA1 and BRCA2) and isolated "these sequences away from other genomic DNA and cellular components." *Id.* at 232. Rather than "creating" BRCA, the court concluded that Myriad merely discovered it. While discovery of the "important correlation" was a "valuable scientific achievement" requiring "technical skill and considerable labor," it was, nevertheless, a "discovery of the handiwork of nature—the natural effect of certain mutations in a particular segment of the human genome." *Id.* at 232. For those reasons, Judge Sweet concluded that despite Myriad's cited structural and functional differences between "native DNA" and "isolated DNA," the two were not "markedly different" from one another. Myriad's composition claims were patent ineligible, as they were directed to natural phenomena.

Judge Sweet next analyzed Myriad's method claims, drawn to 1) analyzing and comparing DNA sequences to determine the existence of BRCA1 and BRCA2; and to 2) screening cancer therapeutics by comparing the growth rate of cells when a test compound is added to one cell group. With the law as it was at the time before the Supreme Court's pronouncements in *Mayo Collaborative Services v. Prometheus Laboratories, Inc.*, 132 S. Ct.

1289 (2012) and *Bilski v. Kappos*, 130 S. Ct. 3218 (2010)¹³ Judge Sweet applied the ‘machine or transformation’ test to find that all of the challenged method claims were drawn to mental processes and lacked any transformative step or machine. *Id.* at 235-37. Myriad argued that its act of “isolating and sequencing DNA” was transformative, and although not explicitly included, should be read into the ‘sequence and analyze’ method claims.

The court concluded that the “preparatory steps” of isolating and sequencing could not be read into the claims—but even if they were, they “would constitute no more than ‘data-gathering step[s]’ that are not ‘central to the purpose of the claimed process.’” *Id.* at 236 (quoting *Bilski v. Kappos*, 545 F.3d 943, 962-63 (Fed. Cir. 2008), *aff’d on other grounds*, *Bilski*, 130 S. Ct. 3218). Likewise, with regard to the ‘comparing growth cell rate’ claims, Judge Sweet concluded the claimed ‘process’ was “in fact, the scientific method itself,” an unpatentable mental process under § 101. *Id.* at 237.

2. Federal Circuit *AMP* Opinions

Myriad appealed Judge Sweet’s decision to the Federal Circuit Court of Appeals. In July 2011, the Federal Circuit affirmed Judge Sweet’s conclusion that Myriad’s “analyze and

¹³ At the time of Judge Sweet’s decision, many courts viewed the machine or transformation test as dispositive when analyzing § 101 challenges to process, or method claims. The Supreme Court in *Bilski* clarified that although the test provides important clues as to patent eligibility, is not categorically dispositive. 130 S. Ct. at 3230. In *Bilski*, the Court considered whether a claimed “process” of risk hedging—reduced in the claims to a mathematical formula—was a patent ineligible abstract idea. The Court noted that business processes may, in some cases, be patentable. Still, it found the formula there at issue to be a patent ineligible abstract idea. *Id.* at 3231. The Court was concerned with allowing a patent on the risk-hedging method based on the formula, where a patent would “pre-empt use of this approach in all fields, and would effectively grant a monopoly over an abstract idea.” *Id.*

compare” method claims were drawn to patent ineligible subject matter, but it reversed on the other issues. 653 F.3d 1329, 1350, 1355-56, 1358 (Fed. Cir. 2011). The Federal Circuit concluded that Myriad’s “isolated DNA” composition claims (including a subset of claims drawn to cDNA) and its method claim covering cancer therapeutic comparisons were patent eligible. *Id.* at 1357-58. The *AMP* plaintiffs successfully petitioned the Supreme Court for a writ of certiorari. 132 S. Ct. 1794 (2012). The Court vacated the Federal Circuit’s decision and remanded the case for further consideration in light of the Supreme Court’s then-recent decision in *Mayo*, 132 S. Ct. 1289.

On remand, the Federal Circuit reached conclusions identical to those in its prior decision. 689 F.3d 1303, 1337 (2012). In an opinion authored by Judge Lourie, two members of the three judge panel concluded that Myriad’s composition claims were drawn to patent eligible subject matter—but the judges did not agree on the reasons for this result. The court again found patent eligible Myriad’s lone method claim relating to comparison of cell growth with or without added therapeutics. *Id.* at 1335-37. Finally, as it had before, the court concluded that Myriad’s method claims for analyzing and comparing DNA sequences were drawn to patent ineligible subject matter. *Id.* at 1334-35.

First, concerning the composition claims, the court cited three representative claims from the ’282 Patent: 1) claim 1, covering “isolated DNA coding for a BRCA1 polypeptide”; 2) claim 2, covering the “isolated DNA of claim 1, wherein said DNA has the nucleotide sequence set forth in” BRCA1 cDNA; and 3) claim 5, covering an “isolated DNA having at least 15 nucleotides of the DNA of claim 1.” *Id.* at 1309. Myriad argued that an isolated DNA molecule claimed in each is “patent eligible because it is . . . ‘a nonnaturally occurring manufacture or composition of matter’ with a distinctive name, character and use.” *Id.* at 1325 (quoting

Myriad's Appellant Brief, 2010 WL 4600106, at *41-42 (other citations omitted)). Myriad asserted that Judge Sweet had 1) mistakenly applied Supreme Court precedent to exclude from patent eligibility all products of nature unless "markedly different" from "naturally occurring" products; and 2) incorrectly focused on the informational content similarity between native DNA and isolated DNA, instead of their claimed differences. Butressing these points, Myriad argued that "isolated DNA does not exist in nature and that isolated DNAs, unlike native DNAs, can be used as primers and probes for diagnosing cancer." *Id.* at 1325.

Plaintiffs disagreed, arguing the isolated DNA Myriad claimed was not markedly different from native DNA, despite Myriad's asserted structural differences, because it "retain[s] part of the same nucleotide sequence" *Id.* at 1326. Further, the plaintiffs warned that if allowed, Myriad's composition claims would "preempt products and laws of nature, excluding anyone from working with the BRCA genes and the genetic information they convey." *Id.*

The United States, as amicus curiae, staked out a "middle ground," arguing that one sub-category of Myriad's claimed isolated DNA, synthetic cDNA, was patent eligible, while the remainder were not. To reach this result, the government proposed viewing the challenged claims through the lens of a so-called 'magic microscope': if one using "an imaginary microscope could focus in on the claimed molecule as it exists in the human body, the claim covers ineligible subject matter." *Id.* at 1326. Applying this test, cDNA would be patent eligible, because it is "engineered by man to splice together non-contiguous coding sequences (i.e., exons)." *Id.* at 1325-26. But "the claimed isolated BRCA1 and BRCA2 sequences" would be patent ineligible, because the nucleotide sequences comprising those genes appear in the body just as they do in the isolated DNA. *Id.* at 1326.

Judges Lourie and Moore found non-cDNA isolated DNA patent eligible, while Judge Bryson concluded it was not. Judge Lourie wrote that all of Myriad's claimed isolated DNA was markedly different from native DNA "in name, character and use." *Id.* at 1328 (quoting *Chakrabarty*, 447 U.S. at 309-310). According to Judge Lourie, the district court was incorrect to focus on the identical nucleotide sequence and informational content in native and isolated DNA because "it is the distinctive nature of DNA molecules as isolated compositions of matter that determines their patent eligibility rather than their physiological use or benefit . . . their informational content is irrelevant." *Id.* In addition to being "free-standing," and "cleaved," isolated DNA may be shorter than native DNA—"synthesized to consist of just a fraction of a naturally occurring DNA molecule"—and thus include "primers and probes, having as few as fifteen nucleotides of a BRCA sequence." *Id.* at 1328.

Writing separately, Judge Moore disagreed with Judge Lourie that isolated DNA—aside from cDNA—was patent eligible simply because of "the chemical differences between genomic and isolated DNA (breaking the covalent bonds)" *Id.* at 1341. Judge Moore concluded that shorter isolated DNA was patent eligible because it possessed the unique utility of being capable of use as primers or probes, and was thus markedly different from naturally occurring DNA. In contrast, Judge Moore expressed concern about the patent eligibility of longer strands of isolated DNA, observing that if she were "deciding [the] case on a blank canvas," she would have found those longer strands, unsuitable for use as primers or probes, patent ineligible. *Id.* at 1343. But, out of deference to the United States Patent and Trademark Office (USPTO) then-existing practice of granting patents on isolated genes and the reliance of those who hold those patents, Judge Moore joined in Judge Lourie's result and concluded that all isolated DNA was patent eligible. *Id.*

Judge Bryson dissented from the panel's conclusion that isolated DNA was patent eligible. *Id.* at 1348. Guided by the Supreme Court's *Chakrabarty* decision, Judge Bryson focused "on two things: 1) the similarity in structure between what is claimed and what it found in nature[;] and 2) the similarity in utility between what is claimed and what is found in nature." *Id.* at 1354. Regarding structure, Judge Bryson concluded that the isolated BRCA genes were not rendered patent eligible simply because they were purified and their bonds cleaved in order to extract them. Like an extracted mineral, a cutting from a plant, or a kidney removed from a body, the only "material change made to these genes from their natural state is the change that is necessarily incidental to [their] extraction" *Id.* at 1350.

That their chemical makeup changes when the bonds are broken does not compel the conclusion that a new molecule is created, as "there is no magic to a chemical bond that requires us to recognize a new product when a chemical bond is created or broken, but not when other atomic or molecular forces are altered." *Id.* at 1349. This is particularly so where Myriad's composition claims, aside from cDNA, are not "defined by any particular chemical formula." *Id.* The fact that isolated DNA has different "terminal groups" than those on naturally occurring genes is insignificant in light of the critical function of DNA—which is "dictated by the nucleotide sequence of the gene—a sequence that is determined by nature and that appears in nature exactly as it appears in the claimed isolated DNA." *Id.* at 1352. Likewise, that the isolated DNA was a smaller part of a naturally larger molecule was of no moment—Judge Bryson found this no more persuasive "than arguing that although an atom may not be patentable, a subatomic particle is patentable because it was previously part of a larger structure." *Id.* at 1353.

In short, Judge Bryson found the structural differences between the isolated BRCA DNA and native genes “irrelevant to the claim limitations, to the functioning of the genes, and to their utility in their isolated form. . . [i]ndeed, that identity of function in the isolated gene is the key to its value.” *Id.* at 1354. The “informational content of the nucleotide sequences is the critical aspect of these molecules” *Id.* at 1355. Thus, the fact that the “nucleotide sequences of the claimed molecules are the same as the nucleotide sequences found in naturally occurring human genes” outweighed the ancillary structural differences. *Id.*

Though they disagreed on isolated non-cDNA DNA, the Federal Circuit panel agreed that cDNA was patent eligible. Judge Lourie analyzed cDNA as a sub-category of isolated DNA, concluding that cDNA was even more distinctive from native DNA than other isolated DNA, as cDNA lacks “the non-coding introns present in naturally occurring chromosomal DNA.” *Id.* at 1329. Judge Moore joined in this part of the opinion. Judge Bryson concurred, generally agreeing that cDNA could be patent eligible as a “human-made invention” with both distinct structure and utility from naturally occurring DNA. *Id.* at 1356 (noting cDNA’s lack of introns and the fact that it can be attached to a promoter and inserted into non-human cell to drive protein expression). But Judge Bryson noted issues with two of Myriad’s claims to short segments of DNA—claims 5 and 6 of the ’282 Patent. Claim 5 covers any segment of DNA defined by claim 1 (isolated DNA coding for BRCA1) at least 15 nucleotides in length, while claim 6 covers any sequence of BRCA1 (exon only) cDNA at least 15 nucleotides long. Judge Bryson noted that claim 6 would encompass “each BRCA1 exon, even though each is naturally defined by transcription.” *Id.* at 1356. Further, the claim would cover exon-only portions of “more than 4% of human genes.” *Id.* Given its breadth, Judge Bryson concluded claim 6 was necessarily drawn to patent ineligible products of nature. *Id.* at 1356.

Though the Federal Circuit found isolated DNA and cDNA patent eligible, it nevertheless found all but one of Myriad's asserted method claims to be patent ineligible as claiming only "abstract mental processes" of "comparing" or "analyzing" BRCA sequences. *Id.* at 1334 (citing *Gottschalk v. Benson*, 409 U.S. 63, 67 (1972)). These claims were not rescued by the fact that the claimed comparisons were limited to the context of BRCA genes or particular alterations. *Id.* (quoting *Bilski*, 130 S. Ct. at 3230 ("prohibition against patenting abstract ideas 'cannot be circumvented by attempting to limit the use of the formula to a particular technological environment.'")) (other citations omitted). Further, like the district court, the Federal Circuit refused to allow Myriad to read into these claims the "additional, allegedly transformative steps" of extracting DNA from a human sample and sequencing the BRCA DNA. *Id.* at 1335. The steps were simply not included in the method claims. *Id.* Rather, the court noted that Myriad's 'compare' and 'analyze' steps were even less transformative than the 'administer' and 'determine' steps in *Mayo*, which were also found to be patent ineligible. *Id.*

Myriad's lone-surviving method claim was claim 20 of the '282 Patent, directed to screening potential cancer therapies by comparing cell growth rates in the presence of different compounds. *Id.* at 1335. Following *Mayo*, the panel found this claim patent eligible where 1) the court had concluded that BRCA genes were patent eligible in the first instance, and 2) it involved a transformative step of growing "host cells transformed with an altered BRCA1 gene." *Id.* at 1336. The court explained that "[t]he transformed, man-made nature of the underlying subject matter in claim 20 makes the claim patent-eligible," and the fact that the claim also included mental steps of determining and comparing growth rates would not alter that conclusion. *Id.*

The *AMP* plaintiffs again appealed their loss on Myriad’s composition claims to the Supreme Court. Myriad did not cross-appeal their invalid method claims. On June 13, 2013, the Court issued a unanimous opinion holding “that a naturally occurring DNA segment is a product of nature and not patent eligible merely because it has been isolated.” 133 S. Ct. at 2111. In contrast, it found that cDNA could be patent eligible “because it is not naturally occurring.” *Id.*

The Court then reviewed the lower court decisions, including the Federal Circuit’s determination “that both isolated DNA and cDNA were patent eligible under § 101.” *Id.* at 2115. Here, the Court noted that Judge Lourie supported his decision that isolated DNA was chemically different from genomic DNA, and therefore patent eligible, by stating, “[i]solated DNA has been cleaved . . . or synthesized to consist of just a fraction of a naturally occurring DNA molecule.” *Id.* at 2115 (quoting 689 F.3d at 1328). The Court also quoted Judge Bryson’s opinion in which he observed that “the structural similarity dwarfs the significance of the structural differences between isolated DNA and naturally occurring DNA, especially where the

structural differences are merely ancillary to the breaking of covalent bonds, a process that is itself not inventive.” *Id.* at 2115 (quoting 689 F.3d at 1355).

Turning to its own analysis of Myriad’s isolated DNA claims, the Court first observed that Myriad did not “create or alter”: 1) “any of the genetic information encoded in the BRCA1 and BRCA2 genes,” as the “location and order of the nucleotides existed in nature before Myriad found them”; or 2) “the genetic structure of DNA.” 133 S. Ct. at 2116. In contrast to inventions with “markedly different characteristics from any found in nature,” the Court identified Myriad’s “contribution” as “uncovering the precise location and genetic sequence of the BRCA1 and BRCA2 genes within chromosomes 17 and 13.” *Id.* (citing *Chakrabarty*, 447 U.S. at 305). While important, useful, and achieved through “extensive research efforts,” this was “not an act of invention.” *Id.* at 2117-19 (noting “extensive effort alone is insufficient to satisfy the demands of § 101” and that the processes used by Myriad to isolate DNA were well understood by geneticists at the time of Myriad’s patents.). Taken together, Myriad’s discovery fell “squarely within the law of nature exception” to patent eligibility under § 101. *Id.*

Critically, the Court found that Myriad’s isolated DNA claims could not be saved “by the fact that isolating DNA from the human genome severs chemical bonds and thereby creates a nonnaturally occurring molecule.” *Id.* This was because Myriad’s claims were “not expressed in terms of chemical composition,” nor did they rely on “chemical changes that result from the isolation of a particular section of DNA.” *Id.* at 2118. Instead, the claims were focused on “the genetic information encoded in the BRCA1 and BRCA2 genes . . . Myriad’s claim is concerned primarily with the information contained in the genetic *sequence*, not with the specific composition of a particular molecule.” *Id.* at 2118 (emphasis in original).

Finally, the Court dismissed Myriad's contention that the USPTO's decision to grant its patents warranted deference. The Court noted that Congress had not only failed to endorse with legislation the type of patents at issue, but also the government had actually taken the position before the Court and the Federal Circuit "that isolated DNA was *not* patent eligible under § 101." *Id.* at 2119 (emphasis in original). Thus, the Court found Myriad's claimed "genes and the information they encode are not patent eligible under § 101 simply because they have been isolated from the surrounding genetic material." *Id.* at 2120.

In contrast, the Court found that cDNA may be patent eligible under § 101. Unlike naturally occurring, isolated DNA segments, "cDNA differs from natural DNA in that 'the non-coding regions [introns] have been removed.'" *Id.* at 2119. A "lab technician unquestionably creates something new when cDNA is made." *Id.* Accordingly, cDNA generally is "distinct from the DNA from which it was derived" and "not a 'product of nature.'" *Id.* But, cDNA's patent eligibility is not without exception. The Court explained that cDNA may be a patent ineligible product of nature when "very short series of DNA may have no intervening introns to remove when creating cDNA." *Id.* In that circumstance, "a short strand of cDNA may be indistinguishable from natural DNA," and presumably not patentable. *Id.* at 2119.

In this respect, the Court's decision echoes Judge Bryson's opinion. Perhaps the Court utilized, without expressly stating, a test similar to the 'magic microscope' urged by the government before the Federal Circuit. The Court's ruling comports with the government's position before the Federal Circuit.

Defendant has released a Cancer Test Requisition Form that offers multiple tests BRCA1 and/or BRCA2 tests. Ford Decl. at Exh. 1. Defendant sequences all coding exons, plus “at least 5 bases into the 5’ and 3’ ends of all the introns and untranslated regions.”¹⁵ Defendant’s tests use primers that amplify the sequences of the exons and at least 20 base pairs of the introns, and the amplicons are sequenced by Next Generation Sequencing. Elliott Decl. at ¶¶ 10, 20-31. Defendant also performs Sanger sequencing as verification of its Next-Generation sequencing to confirm all identified variants or to obtain an additional result if there are parts of the exon that do not have a sufficient number of reads to ensure accuracy. Here, Defendant “sequences the exon or part of the exon in which the variant was detected.”¹⁶

¹⁴ See <<http://ambrygen.com/tests/brcaplus-%E2%80%93-high-risk-breast-cancer-panel>>.

¹⁵ See Defendant's Data Sheets for BreastNext (Dkt. 10 at Exh. D at 4-5); CancerNext (*id.* at Exh. E at 6); OvaNext (*id.* at Exh. F at 5); BRCA1/BRCA2 (*id.* at Exh. I at 2); BRCAplus tests (*id.* at Exh. L at 3); Defendant's NextGen PowerPoint (*id.* at Exh. J at 10-12, 26); RainDance Target Sequencing Assay Manual (*id.* at Exh. K at A-3 to A-7, 3-2, 4-2, 5-2); Elliott Decl. at ¶¶ 10, 20-31.

¹⁶ Elliott 2nd Decl. at ¶ 16; *see also* Elliott Decl. at ¶ 35; Defendant's NGS Cancer Panels PowerPoint (Dkt. 10 at Exh. B at 4); Defendant's Poster (*id.* at Exh. C); Defendant's Data Sheets for BreastNext (*id.* at Exh. D at 5), CancerNext (*id.* at Exh. E at 6), OvaNext (*id.* at Exh. F at 6), BRCA1/2 (*id.* at Exh. I at 2), and BRCAplus tests (*id.* at Exh. L at 3).

Defendant then compares the portions of the patient's genes that have been amplified and sequenced, including the sequence of all of the exonic portions of the BRCA1 and BRCA2 genes as deduced in the sequencing operation, to a wild-type or reference sequence that represents the commonly expected sequence for those genes. The reference sequence that Defendant uses is for the entire "human genome sequence," and thus it aligns all of the sequenced parts of the patient's genes—including the BRCA1 and BRCA2 genes—against the reference human genome sequence and compares the sequenced parts of the patient's genes to that entire reference sequence.¹⁷ Using that comparison, Defendant locates "[a]ll identified variants," removes "variants that have been classified as benign," and performs "Mutation Detection."¹⁸

As additional steps in its BRCA1 and BRCA2 testing services, Defendant's large rearrangement analysis looks for large deletions and duplications of nucleotides. Depending on the test, Defendant uses either: 1) MLPA analysis for "comprehensive (full-gene) gross deletion/duplication analysis"; or 2) microarray analysis to "detect large rearrangement" and identify "gross deletions or duplications in all" genes analyzed.¹⁹ Defendant uses MLPA as part of its BRCA1 and BRCA2 test.²⁰ Defendant uses microarray analysis as part of all of its "panel

¹⁷ Elliott Decl. at ¶¶ 33-34; Defendant's NGS Cancer Panels PowerPoint (Dkt. 10 at Exh. B at 4); Defendant's Data Sheet for BRCA1/BRCA2 Test (*id.* at Exh. I) at 2; Roa Decl. at ¶¶ 27-28.

¹⁸ Elliott Decl. at ¶¶ 33-34; Defendant's NGS Cancer Panels PowerPoint (Dkt. 10 at Exh. B at 4).

¹⁹ Elliott Decl. at ¶¶ 42-50; Defendant's Data Sheet for BRCA1/BRCA2 Test (Dkt. 10 at Exh. I at 2); Defendant's Data Sheets for BRCAplus (*id.* Exh. L at 3); BreastNext (*id.* Exh. D at 5), CancerNext (*id.* at Exh. E at 6); and OvaNext tests (*id.* at Exh. F at 6).

²⁰ Elliott Decl. at ¶¶ 42-49; Defendant's Data Sheet for BRCA1/BRCA2 Test (Dkt. 10 at Exh. I at 2).

tests” that analyze the BRCA1 and BRCA2 genes, including at least its BRCAplus, BreastNext, CancerNext, and OvaNext tests.²¹

Defendant’s MLPA and microarray processes use BRCA1- and BRCA2-specific probes. Specifically, Defendant’s MLPA and microarray processes use “probes targeting the exons or flanking intronic sequences” in the BRCA1 and BRCA2 genes. The probes are complementary to, and thus hybridize to their respective target section if that section is present and has the same nucleotide sequence. Defendant’s probes are specifically targeted for “the wild-type BRCA genes,” and thus are specific for the wild-type alleles.²²

In both the MLPA and microarray processes, Defendant’s BRCA1- and BRCA 2-specific probes are allowed to hybridize to both (1) the synthetic DNA amplicons created from a patient’s sample DNA and (2) the ‘normal’ wild-type reference DNA. Elliott Decl. at ¶¶ 43-48, 50-53; Elliott 2nd Decl. at ¶¶ 12-13; Roa Decl. at ¶¶ 31-37.

In Defendant’s MLPA process, the probes that hybridize to adjacent genomic and wild-type reference sequences are ligated. The polymerase chain reaction process is then applied, amplifying each targeted part of the gene where hybridization of the probes has occurred. If a target section of the gene is present, then the probes hybridize, and the hybridization is detected because the PCR process creates many synthetic DNA molecules having the sequence of the

²¹ Elliott Decl. at ¶¶ 42, 50; Defendant’s Data Sheets for BRCAplus (Dkt. 10 at Exh. L at 3); BreastNext (*id.* at Exh. D at 5); CancerNext (*id.* at Exh. E at 6); and OvaNext tests (*id.* at Exh. F at 6).

²² Elliott Decl. at ¶¶ 43, 47-52; MLPA – an introduction (MRC Holland) (Dkt. 10 at Exh. P) at 1-2; *see* Roa Decl. at ¶¶ 31-32, 35-37. As previously noted, “allele” refers to a form of the gene having a certain sequence, such as variant sequences or mutation sequences of interest and thus may be called a “wild-type allele” or a “mutated allele.” Roa Decl. at ¶ 31.

target section for that probe. If a target section is not present, then the probes for that section do not hybridize to anything, and the fact that the target section is missing will be evident from the relative lack of synthetic DNA molecules corresponding to that section made during the PCR process. Elliott Decl. at ¶¶ 43-48; *see* Roa Decl. at ¶¶ 31-34. The resulting data allows Defendant to compare the patient's germline sequence to a wild-type sequence. Elliott Decl. at ¶¶ 46-48; Roa Decl. at ¶ 34. The data resulting from Defendant's microarray analysis identify both the existence of the allele and the existence of large base pair mutations in a germline nucleic acid sequence by comparing that sequence to the wild-type allele. Elliott Decl. at ¶¶ 50-53; *see id.* at ¶¶ 43-48; *see* Roa Decl. at ¶¶ 35-37.

G. The Patent Claims at Issue in Plaintiffs' Motions for Preliminary Injunction

In July 2013, Plaintiffs moved for a preliminary injunction seeking to prohibit Defendant from selling testing services based on alleged infringement of Myriad's patents. In their Motion for Preliminary Injunction, Plaintiffs allege that they are likely to prove that Defendant is infringing:

- 1) claims 16 and 17 of the '282 Patent;
- 2) claims 29 and 30 of the '492 Patent;
- 3) claims 7 and 8 of the '441 Patent;
- 4) claim 4 of the '857 Patent;
- 5) claim 5 of the '721 Patent; and
- 6) claims 2 and 4 of the '155 Patent.

(Dkt. 5 at 15-30.)

The allegedly infringed claims may be divided into two general categories. The first covers claims drawn to compositions of matter, and are collectively referred to as the Primer

Claims. These Primer Claims include claims 16 and 17 of the '282 Patent, and claims 29 and 30 of the '492 Patent. The remaining claims cover testing processes relating to BRCA1 and BRCA2, and the court collectively refers to them as the Method Claims. Each of the Method Claims is drawn to the mental process of comparing a genomic DNA sample to a DNA sequence that may be found in the BRCA1 and BRCA2 genes.

1. The Primer Claims

Claim 16 of the '282 Patent provides:

16. A pair of single-stranded DNA primers for determination of a nucleotide sequence of a BRCA1 gene by a polymerase chain reaction the sequence of said primers being derived from human chromosome 17q wherein the use of said primers in a polymerase chain reaction results in the synthesis of DNA having all or part of the sequence of the BRCA1 gene.

Claim 17 of the '282 Patent provides:

17. The pair of primers of claim 16 wherein said BRCA1 gene has the nucleotide sequence set forth in SEQ ID NO:1.

The term "SEQ ID NO:1" refers to the nucleotide sequence of the BRCA1 cDNA—the exon-only nucleotide sequence of the BRCA1 gene—as depicted in the Sequence Listing of the '473 Patent, the '282 Patent, and the '999 Patent. '473 Patent col.52 ll.50-56; '282 Patent col.53 ll.4-9; and '999 Patent col.53 ll.16-22; Kay Decl. from *AMP* Litigation at ¶ 49 (Dkt. 34-4).

Claim 29 of the '282 Patent provides:

29. A pair of single-stranded DNA primers of at least 15 nucleotides in length for determination of the nucleotide sequence of a BRCA2 gene by a polymerase chain reaction, the sequence of said primers being isolated from human chromosome 13, wherein the use of said primers in a polymerase chain reaction results in the synthesis of DNA comprising all or at least 15 contiguous nucleotides of the BRCA2 gene.

Claim 30 of the '282 Patent provides:

30. The pair of primers of claim 29 wherein said BRCA2 gene has the nucleotide sequence set forth in SEQ ID NO:1.

In the context of the BRCA2 patents, the term “SEQ ID NO:1” refers to the nucleotide sequence of the BRCA2 cDNA—the exon-only sequence—as depicted in the Sequence Listing of the ’492 Patent. ’492 Patent col.44 l.53—col.45 l.10; Kay Decl. from *AMP* Litigation at ¶ 50.

2. The Method Claims

a. Claims 7 and 8 of the ’441 Patent

Claims 7 and 8 of the ’441 Patent depend on claim 1, which is set forth below. The Federal Circuit held in the *AMP* litigation that claim 1 is drawn to patent ineligible subject matter—mental processes of comparing. The patent ineligible claim 1 recites:

1. A method for screening germline of a human subject for an alteration of a BRCA1 gene which comprises comparing germline sequence of a BRCA1 gene or BRCA1 RNA from a tissue sample from said subject or a sequence of BRCA1 cDNA made from mRNA from said sample with germline sequences of wild-type BRCA1 gene, wild-type BRCA1 RNA or wild-type BRCA1 cDNA, wherein a difference in the sequence of the BRCA1 gene, BRCA1 RNA or BRCA1 cDNA of the subject from wild-type indicates an alteration in the BRCA1 gene in said subject.

Claim 7 of the ’441 Patent is drawn to the method of the patent ineligible claim 1, but

specifically requires the use of probes:²³

7. The method of claim 1 wherein a germline nucleic acid sequence is compared by hybridizing a BRCA1 gene probe which specifically hybridizes to a BRCA1 allele to genomic DNA isolated from said sample and detecting the presence of a hybridization product wherein a presence of said product indicates the presence of said allele in the subject.

Claim 8 is drawn to the method of claim 1, but more specifically with the use of primers:

“The method of claim 1 wherein a germline nucleic acid sequence is compared by amplifying all

²³ At column 21, lines 35-50, the specification of the '441 Patent defines “probes” used in detecting BRCA1 alleles which predispose to certain cancers as being specific for those alleles:

“Probes”. Polynucleotide polymorphisms associated with BRCA1 alleles which predispose to certain cancers or are associated with most cancers are detected by hybridization with a polynucleotide probe which forms a stable hybrid with that of the target sequence, under stringent to moderately stringent hybridization and wash conditions. If it is expected that the probes will be perfectly complementary to the target sequence, stringent conditions will be used. Hybridization stringency may be lessened if some mismatching is expected, for example, if variants are expected with the result that the probe will not be completely complementary. Conditions are chosen which rule out nonspecific/adventitious bindings, that is, which minimize noise. Since such indications identify neutral DNA polymorphisms as well as mutations, these indications need further analysis to demonstrate detection of a BRCA1 susceptibility allele.

At column 15, lines 29-43, the specification of the '441 Patent teaches that probes used to identify mutations are “allele-specific,” that is, are specific for a particular mutation:

DNA sequences of the BRCA1 gene which have been amplified by screened [sic] using allele-specific probes. These probes are nucleic acid oligomers, each of which contains a region of the BRCA1 gene sequence harboring a known mutation. For example, one oligomer may be about 30 nucleotides in length, corresponding to a portion of the BRCA1 gene sequence. By use of a battery of such allele-specific probes, PCR amplification products can be screened to identify the presence of a previously identified mutation in the BRCA1 gene. Hybridization of allele-specific probes with amplified BRCA1 sequences can be performed, for example, on a nylon filter. Hybridization to a particular probe under stringent hybridization conditions indicates the presence of the same mutation in the tumor tissue as in the allele-specific probe.

or part of a BRCA1 gene from said sample using a set of primers to produce amplified nucleic acids and sequencing the amplified nucleic acids.”

b. Claim 4 of the '857 Patent

The Federal Circuit concluded in its *AMP* decision that claim 2 of the '857 Patent, on which claim 4 depends, is drawn to patent ineligible subject matter—abstract mental processes.

Claim 2 provides:

2. A method for diagnosing a predisposition for breast cancer in a human subject which comprises comparing the germline sequence of the BRCA2 gene or the sequence of its mRNA in a tissue sample from said subject with the germline sequence of the wild-type BRCA2 gene or the sequence of its mRNA, wherein an alteration in the germline sequence of the BRCA2 gene or the sequence of its mRNA of the subject indicates a predisposition to said cancer.

Claim 4 of the '857 Patent recites the method of claim 2, but adds the use of an assay:

4. The method of claim 2 wherein the detection in the alteration in the germline sequence is determined by an assay selected from the group consisting of (a) observing shifts in electrophoretic mobility of single-stranded DNA on non-denaturing polyacrylamide gels, (b) hybridizing a BRCA2 gene probe to genomic DNA isolated from said tissue sample, (c) hybridizing an allele-specific probe to genomic DNA of the tissue sample, (d) amplifying all or part of the BRCA2 gene from said tissue sample to produce an amplified sequence and sequencing the amplified sequence, (e) amplifying all or part of the BRCA2 gene from said tissue sample using primers for a specific BRCA2 mutant allele, (f) molecularly cloning all or part of the BRCA2 gene from said tissue sample to produce a cloned sequence and sequencing the cloned sequence, (g) identifying a mismatch between (1) a BRCA2 gene or a BRCA2 mRNA isolated from said tissue sample, and (2) a nucleic acid probe complementary to the human wild-type BRCA2 gene sequence, when molecules (1) and (2) are hybridized to each other to form a duplex, (h) amplification of BRCA2 gene sequences in said tissue sample and hybridization of the amplified sequences to nucleic acid probes which comprise wild-type BRCA2 gene sequences, (i) amplification of BRCA2 gene sequences in said tissue sample and hybridization of the amplified sequences to nucleic acid probes which comprise mutant BRCA2 gene sequences, (j) screening for a deletion mutation in said tissue sample, (k) screening for a point mutation in said tissue sample, (l) screening for an insertion mutation in said tissue sample, (m) in situ hybridization of the BRCA2 gene of said tissue sample with nucleic acid probes which comprise the BRCA2 gene.

c. Claim 5 of the '721 Patent

Claim 5 of the '721 Patent depends on claim 1, which recites:

5. A method for determining an omi haplotype of a human BRCA1 gene comprising:

- (a) determining the nucleotide sequence of the BRCA1 gene or fragment thereof from at least one female individual with a family history which indicates a predisposition to breast cancer,
- (b) comparing the determined nucleotide sequence from said female individual to SEQ ID NO: 263,²⁴ and
- (c) determining the presence of the following nucleotide variations: thymine at nucleotides 2201 and 2731, cytosine at nucleotides 2430 and 4427, and guanine at nucleotides 3232, 3667 and 4956, wherein the presence of the nucleotide variations in the determined nucleotide sequence indicates the omi1 haplotype.

Claim 5 of the '721 Patent recites, "The method of claim 1 wherein the BRCA1 gene or fragment thereof is amplified prior to nucleotide sequencing."

d. Claims 2 and 4 of the '155 Patent

Claim 2 of the '155 Patent recites:

2. A method of identifying individuals having a BRCA1 gene with a BRCA1 coding sequence not associated with breast or ovarian cancer comprising:

- a) amplifying a DNA fragment of an individual's BRCA1 coding sequence using an oligonucleotide primer which specifically hybridizes to sequences within the gene; b) sequencing said amplified fragment by dideoxy sequencing; c) repeating steps (a) and (b) until said individual's BRCA1 coding sequence is completely sequenced; d) comparing the sequence of said amplified DNA to the sequence of SEQ. ID. NO: 1; e) determining the presence or absence of each of the following polymorphic variations in said individual's BRCA1 coding sequence:

AGC and AGT at position 2201,

TTG and CTG at position 2430, CCG and CTG at position 2731,

²⁴ "SEQ ID NO:263" in claim 1 of the '721 Patent refers to a contiguous cDNA sequence of all of the exons of BRCA1. *See* '721 Patent cols.109-121; Pls.' Reply Br. at 62-63.

GAA and GGA at position 3232,

AAA and AGA at position 3667,

TCT and TCC at position 4427, and AGT and GGT at position 4956; f) determining any sequence differences between said individual's BRCA1 coding sequences and SEQ. ID. NO: 1 wherein the presence of any of the said polymorphic variations and the absence of a polymorphism outside of positions 2201, 2430, 2731, 3232, 3667, 4427, and 4956, is correlated with an absence of increased genetic susceptibility to breast or ovarian cancer resulting from a BRCA1 mutation in the BRCA1 coding sequence.

Claim 4 of the '155 Patent recites:

4. A method of detecting an increased genetic susceptibility to breast and ovarian cancer in an individual resulting from the presence of a mutation in the BRCA1 coding sequence, comprising:

a) amplifying a DNA fragment of an individual's BRCA1 coding sequence using an oligonucleotide primer which specifically hybridizes to sequences within the gene;

b) sequencing said amplified fragment by dideoxy sequencing;

c) repeating steps (a) and (b) until said individual's BRCA1 coding sequence is completely sequenced;

d) comparing the sequence of said amplified DNA to the sequence of SEQ. ID. NO: 1;

e) determining any sequence differences between said individual's BRCA1 coding sequences and SEQ. ID. NO: 1 to determine the presence or absence of polymorphisms in said individual's BRCA coding sequences wherein a polymorphism which is not any of the following: AGC or AGT at position 2201, TTG or CTG at position 2430, CCG or CTG at position 2731, GAA or GGA at position 3232, AAA or AGA at position 3667, TCT or TCC at position 4427, and AGT or GGT at position 4956;

is correlated with the potential of increased genetic susceptibility to breast or ovarian cancer resulting from a BRCA1 mutation in the BRCA1 coding sequence.

"SEQ ID NO:1" recited in claims 2 and 4 of the '155 Patent is a contiguous cDNA sequence of all of the exons of BRCA1. '155 Patent col.19 ("Molecule Type: cDNA"); Pls.' Reply Br. at 62-63.

II. DISCUSSION

A. Preliminary Injunction Standards

The court “may grant injunctions in accordance with the principles of equity to prevent the violation of any right secured by patent, on such terms as the court deems reasonable.” 35 U.S.C. § 283. The Federal Circuit cautions, however, that “[a] preliminary injunction is a ‘drastic and extraordinary remedy that is not to be routinely granted.’” *National Steel Car, Ltd. v. Canadian Pac. Ry., Ltd.*, 357 F.3d 1319, 1324 (Fed. Cir. 2004) (quoting *Intel. Corp. v. ULSI Sys. Tech., Inc.*, 995 F.2d 1566, 1568 (Fed. Cir. 1993)).

To obtain this extraordinary remedy, Plaintiffs must show four factors:²⁵

A plaintiff seeking a preliminary injunction must establish that [it] is likely to succeed on the merits, that [it] is likely to suffer irreparable harm in the absence of preliminary relief, that the balance of equities tips in [its] favor,²⁶ and that an injunction is in the public interest.

AstraZeneca, L.P. v. Apotex, Inc., 633 F.3d 1042, 1049 (Fed. Cir. 2010) (rehearing denied en banc Jan. 31, 2011) (quoting *Winter v. Natural Res. Def. Council, Inc.*, 555 U.S. 7, 20 (2008)); see also *Aria Diagnostics v. Sequenom, Inc.*, 726 F.3d 1296, 1304 (Fed. Cir. 2013) (noting that district court “correctly held that in addition to showing the likelihood of success on the merits,

²⁵ The parties disagree about whether Federal Circuit or Tenth Circuit preliminary injunction standards apply to the preliminary injunction analysis in this case, with Defendant arguing that the regional circuit standards apply. The Federal Circuit has clearly stated that “a preliminary injunction enjoining patent infringement pursuant to 35 U.S.C. § 283 ‘involves substantive matters unique to patent law and, therefore, is governed by the law of this court.’” *Revision Military, Inc. v. Balboa Mfg. Co.*, 700 F.3d 524, 525 (Fed. Cir. 2012) (quoting *Hybritech Inc. v. Abbott Labs*, 849 F.2d 1446, 1451, n.12 (Fed. Cir. 1988)). Since the “issuance of an injunction pursuant to this section enjoins the violation of any right secured by patent . . . a preliminary injunction of this type, although a procedural matter, involves substantive matters unique to patent law and, therefore, is governed by the law of this court.” *Id.* (quoting *Hybritech*, 849 F.2d at 525) (other citations omitted). The court applies Federal Circuit law here.

²⁶ This factor is often discussed in terms of the balance of the parties’ comparative hardships. *Abbott Labs. v. Sandoz, Inc.*, 544 F.3d 1341, 1344 (Fed. Cir. 2008) (citations omitted).

The Federal Circuit cautions courts not to assume that a showing of irreparable harm is inadequate merely because patentees may eventually recover economic damages for infringement. That assumption strips patents of “their character as an exclusive right as articulated by the Constitution” and causes them to “become at best a judicially imposed and monitored compulsory license.” *Id.* at 1304. Further, loss of market share can constitute irreparable harm even if it only moderately affects a patentee company’s profitability or a small part of a company’s business. *Robert Bosch, LLC v. Pylon Mfg. Corp.*, 659 F.3d 1142, 1152 (Fed. Cir. 2011) (quoting *Hoffman-LaRoche, Inc. v. Cobalt Pharm. Inc.*, 2010 WL 4687839, at *12 (D.N.J. Nov. 10, 2010) (a patentee’s “size and profitability, and the small impact the likely harms would have on [its] overall profitability . . . says nothing about whether such harms are irreparable.”)).

- 1) through price erosion for Myriad's testing products, causing Myriad to lose the benefit of its established pricing strategy;
- 2) by reducing Myriad's testing product market share;
- 3) reputational injury to Myriad because the public will mistakenly associate flawed test results from the Defendant's testing with Myriad's allegedly superior testing products, resulting in dilution of Myriad's brand; and

- 4) by simply losing the benefit of the remainder of Myriad's patent terms, including Plaintiffs' right to exclude competitors, particularly where some of the patents at issue will begin to expire in August 2014.

The court discusses these alleged harms in turn.

1. Price Erosion and Loss of Market Share

For seventeen years, Myriad was the only company in the United States offering a full sequence test for the BRCA1 and BRCA2 genes. Ford Decl. at ¶ 8. That changed in June 2013, when Defendant began offering a multi-gene panel test for BRCA1 and BRCA2 at \$2,200, significantly below Myriad's \$4,040 integrated BRCAAnalysis test (including BART). Ford Decl. at ¶ 11; Def's. Opp. Memo. at 15, ¶ 46.

Plaintiffs and their economist expert, Dr. James R. Kearl, claim that Plaintiffs are likely to suffer irreparable harm absent an injunction because Myriad is likely to lose its share of the BRCA1 and BRCA2 testing market—a market it previously monopolized in the United States. Plaintiffs contend that those seeking Defendant's lower priced testing likely represent lost Myriad customers. Further, Plaintiffs allege that Myriad faces a substantial threat of losing its third-party payor customers unless it lowers prices, thus leading to erosion of its pricing structure. Simply put, in a BRCA testing market where Myriad had been the lone seller, the

introduction of new competitors offering alternative testing will force Myriad to choose between lowering its price or losing customers.²⁷

Defendant, relying on its CPA expert, Scott Hampton, offers a number of arguments in response. First, Defendant points out that Myriad has never before dropped its testing prices in response to public pressure, and suggests that this pricing stubbornness shows that Myriad will not now lower its prices, even in the face of competition. Hampton Decl. at ¶ 23. The court finds this entirely unpersuasive. That Myriad did not respond to public pressure says nothing about Myriad's response to new competitors entering its previously exclusive market with competing, lower-priced testing products. It is clear that pressure on Myriad to lower prices will be more acute and more effective than when customers had no market alternative. Regardless, if Myriad does not submit to that pressure to lower prices, it will undoubtedly lose market share. Kearl Decl. at 3, 7, 16-18.

Second, Defendant contends that the BRCA testing market is largely controlled by decision makers within third-party payors, such as insurers and health maintenance organizations. These third-party payors decide whether to pay for or reimburse testing costs. Defendant points out that Myriad has fixed long-term pricing agreements with many of these third-party payors. Thus, Defendant claims, Myriad will not need to lower its prices anytime soon. Hampton 2nd Decl. at ¶ 6 (Dkt. 134). Defendant also argues Plaintiffs have yet to identify

²⁷ There is an inverse relationship between the potential for price erosion and loss of market share. As Dr. Kearl explains, "to the degree that Myriad lowers its price in response to competition" from other testing companies, the price differences between Myriad and those companies "will be less, and the Myriad [market] share loss will also be less. However, in this case, while market share losses will be smaller, price erosion losses will be higher." Kearl Decl. at 17, n.19.

any third-party payor that has cancelled an agreement or successfully re-negotiated an existing long-term contract to obtain a lower testing price.

But Defendant fails to recognize that Myriad permits those third-party payors to cancel their long-term contracts with 90 days' notice to Myriad. Sept. 12, 2013 Hearing Tr. (Ford) at 314:18-24. Moreover, Plaintiffs recently submitted evidence that at least one third-party payor, the Centers for Medicare and Medicaid Services (CMS), has decided to lower its reimbursement rate for BRCA1 and BRCA2 testing from \$2,700 to \$1,438.14. (Dkt. 177-1.)²⁸ The evidence before the court shows that Myriad is clearly experiencing pressure from third-party payors to reduce prices. Ford Decl. at ¶ 16. Notwithstanding that CMS is the lone third-party payor identified by Plaintiffs as lowering reimbursement rates, and that Plaintiffs have thus far avoided any other renegotiated or cancelled contracts, Plaintiffs have shown that it is only a matter of time before the pricing pressure from Defendant's testing causes renegotiated or cancelled contracts. Without an injunction, it is clear that Myriad will have to either lower its testing price or lose market share—or possibly both.

Third, Defendant contends that an injunction is unnecessary even if market forces cause Myriad to reduce its prices because Myriad will be able to raise prices again if Plaintiffs ultimately prevail in this lawsuit on the merits. Hampton Decl. at ¶ 25. But Dr. Kearl persuasively notes that once prices drop, Myriad will face daunting resistance to reinstating higher prices. Kearl Decl. at 11-13; *see also Sanofi-Synthelabo v. Apotex, Inc.*, 470 F.3d 1368, 1382 (Fed. Cir. 2006) (*reh 'g and reh 'g en banc* denied Jan. 19, 2007) (affirming entry of

²⁸ The court has taken judicial notice that CMS lowered its reimbursement rate on January 1, 2014. (Dkt. Nos. 177-1 and 181.) This rate remains subject to change following closure of a public comment period, originally slated for January 27, 2014, but which has been extended to February 28, 2014. *See id.*; and *Gapfill Pricing Inquiries*, <[http:// www. cms.gov/ Medicare/ Medicare-Fee-for- Service-Payment/ClinicalLabFeeSched/Gapfill-Pricing-Inquiries.html](http://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/ClinicalLabFeeSched/Gapfill-Pricing-Inquiries.html)>.

preliminary injunction for drug patentee; citing with approval district court's irreparable harm analysis which found that it would be nearly impossible to restore pre-competition pricing). The court finds Dr. Kearl's price-resistance analysis persuasive, particularly where Myriad's patents will begin to expire in August 2014, long before any resolution of this dispute on the merits. Kearl Decl. at 11-15; Ford Decl. at ¶ 21; *see also Pfizer, Inc. v. Teva Pharm., USA, Inc.*, 429 F.3d 1364, 1380 (Fed. Cir. 2005) (noting that district court properly concluded that accused drug patent infringer's sales of generic drug "would cause substantial harm to [the patentee] and loss of statutory right to exclude . . . for the [relatively short] remaining life of the . . . patent . . .").

Fourth, Defendant appears to argue that Myriad has the financial capability to keep its pricing structure, even if it causes Myriad to lose money or market share. *See* Def's. Opp. Memo. at 89 (noting that Myriad's "\$400 million in cash and cash equivalents will allow it to maintain its current pricing structure during the pendency of this suit if it chooses to do so."); Hampton Decl. at ¶ 66. Defendant also points out that Myriad has projected continued revenue growth for Fiscal Year 2014, despite competition from BRCA1 and BRCA2 testing companies. Hampton Decl. at ¶¶ 52-53. Based on these projections, Defendant contends that Myriad cannot seriously believe it will suffer irreparable injury, even if it experiences price erosion or loss of market share.

Defendant's argument fails to rebut Dr. Kearl's logical and persuasive testimony that Plaintiffs will suffer irreparable financial harm if an injunction does not issue. This harm need not destroy Myriad in order to be irreparable. The Federal Circuit has explained that a loss of market share can constitute irreparable harm even if it only moderately affects profitability or a portion of a patentee's business. *Robert Bosch*, 659 F.3d at 1152 (the "fact that an infringer's harm affects only a portion of a patentee's business says nothing about whether that harm can be

rectified.”) (citing *Hoffman-LaRoche, Inc.*, 2010 WL 4687839, at *12 (rejecting notion that harm is not irreparable if patentee is large and profitable, and infringement allegedly has “small impact . . . on . . . overall profitability.”)). That Myriad continues to project overall profitability does not “automatically rebut a case for irreparable injury.” *Douglas Dynamics*, 717 F.3d at 1344 (rejecting district court’s reasoning that permanent injunction should not issue where patentee snow plow company’s market share increased by one percent per year after accused infringer introduced its competing snow plow assembly; noting market share might increase for various reasons unrelated to the infringing conduct).

Finally, Defendant contends that its entry into the BRCA testing market has actually expanded that market, particularly for those who seek less expensive tests and “meaningful second opinion testing.” Def’s. Opp. Memo. at 95. Defendant argues Myriad cannot be harmed “by the loss of sales it never would have realized.” *Id.* at 96. Defendant does not contend that complete mutual exclusivity exists between potential Myriad customers, whether third-party payors or individuals, and those who will seek out Defendant’s less expensive tests during the remainder of Myriad’s patent terms. It defies reason to conclude that Defendant’s customers will comprise only persons who could not have obtained via direct payment or a third-party payor a higher-priced Myriad BRCA1 and BRCA2 test. Nor does the record in any way support such a finding.

The court concludes that Plaintiffs have adequately demonstrated they are likely to suffer irreparable harm in the form of price erosion and loss of market share if an injunction does not issue.

would offer time to educate the public about the superiority of their testing—including the newly introduced myRisk test. Ford Decl. at ¶¶ 17, 21-23.

Plaintiffs' arguments here fall short. First, there is no competent evidence showing that Defendant's tests are meaningfully less accurate or reliable than Myriad's tests. Defendant compellingly points out that Myriad's VUS rate is based on data that Myriad does not share with the public and is difficult to verify. Defendant's current VUS rates are 4.2 percent. Chao Decl. at ¶¶ 52-53; Chao 2nd Decl. ¶ 2 (Dkt. 137).³⁰ These rates are already low and will go down over time. VUS rates are highest right after a test is first offered, but they decrease as more tests are run and more data becomes available to assist with classification. Swisher Decl. at ¶¶ 52-53; Chao Decl. at ¶ 52. Moreover, Plaintiffs simply provide no evidence of consumer confusion between testing companies.

But Plaintiffs insist they may suffer reputational harm if Defendant's entry into the BRCA testing market causes "dilution of Myriad's reputation," where Myriad was until recently "the only provider of genetic predisposition testing for breast and ovarian cancer . . . and thus providers and patients naturally associate the test with Myriad." Pls.' Reply Br. at 136. In *Douglas Dynamics*, the plaintiff snow plow company sought a permanent injunction against a competing snow plow company. 717 F.3d at 1338-39. The district court denied the injunction, stating there was no reputational harm because there was no customer confusion. *Id.* at 1344. The Federal Circuit disagreed, finding that reputational harm may result even without actual

³⁰ In their Reply Brief, Plaintiffs express skepticism about this rate but do not dispute it with any evidence. Geier Decl. at ¶ 29 (Dkt. 104) (noting that Defendant "is reportedly quoting a 5% VUS rate, but I don't believe that estimate has yet been validated since they have only been doing testing for a few weeks."). Defendant notes that while Plaintiffs question its VUS rate, it is actually more transparent in its VUS rate determination than Myriad is because Defendant provides full variant data information to patients, their doctors, and genetic counselors through test reports. Chao Decl. at ¶¶ 58-59, Exh. L; Swisher Decl. at ¶¶ 46-49, Exhs. G, I.

customer confusion. This was due to the possibility that consumers would see the patentee's innovations appearing in products "considered less prestigious and innovative," or that the patentee's dealers might believe it "did not enforce its intellectual property rights." *Id.*

In contrast, Plaintiffs here offer no clear evidence suggesting that the public would view Defendant's testing products as less prestigious or innovative. In fact, Myriad's newly launched myRisk test is a multi-gene panel test that utilizes Next Generation Sequencing—much like Defendant's multi-panel CancerNext test.³¹ Plaintiffs' contention that their reputations will be besmirched by the continued sale of Defendant's testing products does not withstand scrutiny. At least at this preliminary stage, the court concludes that Plaintiffs have not shown that Defendant's testing is likely to cause irreparable harm to Plaintiffs' reputations.

3. Loss of Remainder of the Exclusive Patent Terms

Plaintiffs also contend that Defendant's alleged infringement deprives Plaintiffs of enjoyment of their patents' exclusive terms. Myriad says this causes irreparable harm because its long-term corporate strategy was based on an expectation that Plaintiffs would enjoy their exclusive patent terms until they began to expire in August 2014, giving Myriad time to develop and introduce to the public new products, including the recently released myRisk cancer test. *Robert Bosch*, 659 F.3d at 1149 (citing *Acumed LLC v. Stryker Corp.*, 551 F.3d 1323, 1328 (Fed. Cir. 2008) (in view of right to exclude, "infringement may cause a patentee irreparable harm not remediable by a reasonable royalty.")).

Defendant responds that because Myriad has not yet sued every company that is or has announced plans to begin BRCA1 and BRCA2 testing, Myriad has inconsistently enforced its patents. It claims that this demonstrated Plaintiffs' indifference to their patents' exclusive terms.

³¹ Myriad launched myRisk on September 5, 2013.

But the Federal Circuit has observed that “[p]icking off one infringer at a time is not inconsistent with being irreparably harmed.” *Robert Bosch, LLC*, 659 F.3d at 1151 (quoting *Pfizer, Inc. v. Teva Pharm. USA, Inc.*, 429 F.3d at 1381). Were the court to find otherwise, it would “effectively establish a presumption against irreparable harm whenever the market contains a plurality of players”:

Under such circumstances, the first infringer sued could always point to the existence of additional competitors. And, perversely, if that infringer were to succeed in defeating an injunction, subsequent adjudged infringers could point to the market presence of the first infringer when opposing a request for an injunction. Consequently, without additional facts showing that the presence of additional competitors renders the infringer's harm reparable, the absence of a two-supplier market does not weigh against a finding of irreparable harm.

Id. (noting that patent holder had “diligently pursued infringers”; reversing district court’s finding that the “absence of a two-player market effectively prohibit[ed] a finding of irreparable harm.”). Thus, even if Plaintiffs have not sued every company that has announced an intent to conduct BRCA1 and BRCA2 testing, this does not mean that Plaintiffs cannot show irreparable harm here.

Moreover, it appears that Plaintiffs have been diligent, and not indifferent, in enforcing their patents through litigation in the recent aftermath of the Supreme Court’s decision in *AMP* just a few months ago. *See supra* note 3 (discussing current litigation involving Plaintiffs and alleged infringers). By all indications, Myriad and the other Plaintiffs are actively defending their patent exclusivity. The court finds that the Plaintiffs’ claimed harm from the loss of exclusivity of Myriad’s patent terms bolsters Plaintiffs’ showing that they likely will suffer irreparable harm without injunctive relief.

4. Damages Calculation and Defendant's Ability to Pay

Defendant claims that even if Plaintiffs have shown a likelihood of harm, particularly through price erosion, this harm is compensable with money damages, rendering a preliminary injunction unwarranted. Defendant relies on its CPA expert, Scott Hampton, who states that because "Myriad enjoyed a near perfect monopoly prior to June of 2013," there will be a clean starting point from which to calculate Plaintiffs' damages. Hampton Decl. at ¶ 31. In considering this issue, the court must heed the clear instruction from the Federal Circuit to act with caution in assuming that money damages will suffice when Plaintiffs have shown harms likely flowing from price erosion, loss of market share, and loss of patent terms, as is the case here. *Aria Diagnostics*, 726 F.3d at 1304.

Plaintiffs maintain that unless an injunction issues, the complex pricing and sales factors in this case present a substantial danger that they will be undercompensated if they prevail on the merits. Kearl Decl. at 15. The court agrees.

First, Myriad does not charge one set price for every test, but instead may alter its pricing by test type, customer, and time period. Kearl Dec. at 8. It will be especially challenging here to determine, once all prices decline, the price that customers would have paid Myriad but for Defendant's entry into the market.

Second, Defendant appears to assume that if Plaintiffs eventually prevail, Myriad will be able to reverse the price erosion that will occur without an injunction. As explained above, the court agrees with Plaintiff's economist, Dr. Kearl, who concludes that this is unlikely. Kearl Decl. at 11-13.

Finally, in the aftermath of the Supreme Court's *AMP* decision, numerous companies are offering or planning to offer close substitute BRCA1 and BRCA2 testing. Although Myriad has

been seeking to stop the accused infringers through litigation, it will be very challenging for Plaintiffs to show the degree to which any accused infringer should be held responsible for price erosion. Or, if Myriad does not lower its pricing, it will be difficult to show which sales a competitor enjoys would otherwise have been Myriad's, and at what price. *Kearl Decl.* at 15-16.

As Dr. Kearl states:

When the damages estimation becomes more complex, the losses that are clearly the result of the specific challenged conduct of the specific defendant before the court become a smaller amount. In this way, complexity tends to lead—in practice—to lower economic damages estimates and awards, and under compensation of the plaintiff.

Id. at 18.

Moreover, the court is concerned with Defendant's ability to pay a large damage award if Plaintiffs prevail in this litigation. The Federal Circuit has found that an accused infringer's "lack of financial wherewithal to satisfy a judgment" may be considered in evaluating irreparable harm. *Robert Bosch*, 659 F.3d at 1151. Here, Defendant and its experts allege that Defendant will be forced to close its doors and terminate its 180 employees if a preliminary injunction is entered. This is due to substantial investments made to prepare to conduct BRCA1 and BRCA2 testing and the centrality of that testing to its revenue stream. *Hampton Decl.* at ¶ 57. This allegation raises serious concerns about Defendant's ability to satisfy a substantial damage award.

For these reasons, the court concludes that Plaintiffs have shown they are likely to suffer irreparable harm in the form of lost market share, price erosion, and loss of the remainder of their exclusive patent term if a preliminary injunction does not issue.

patent because the USPTO “rejects claims if they are drawn to ineligible subject matter, just as it rejects claims if not compliant with §§ 102, 103, or 112.”³² *Id.*

An accused infringer attacking a patent’s validity at the preliminary injunction stage bears the burden to “come forward with evidence of invalidity, just as it would be at trial.” *Titan Tire*, 566 F.3d at 1377. The patentee must then respond with contrary evidence to show a likelihood of success on the merits. *Id.* At this early stage, however, the court “does not resolve the validity question,” but instead assesses “the persuasiveness of the challenger’s evidence, recognizing that it is doing so without all the evidence that may come out at trial.” *Id.* (citations omitted). In other words, the alleged infringer need not persuade the court that the patent is invalid. Rather, in seeking the extraordinary relief of a preliminary injunction, “it is the patentee, the movant, who must persuade the court that, despite the challenge presented to validity, the patentee nevertheless is likely to success at trial on the validity issue.” *Id.*

With regard to the presumption of subject matter eligibility in the area of gene patents, the Supreme Court in *AMP* notably rejected Myriad’s contention that the USPTO’s “past practice of awarding gene patents is entitled to deference.” 133 S. Ct. at 2118 (citations omitted). The Court pointed out that Congress had not endorsed with subsequent legislation the USPTO’s practice in this area. *Id.* Further, in the *AMP* litigation, the United States had argued “that isolated DNA was *not* patent eligible under § 101,” and that the USPTO’s practice was not ‘a sufficient reason to hold that isolated DNA is patent-eligible.’” *Id.* (emphasis in original)

³² Myriad claims that because the *AMP* plaintiffs did not challenge every claim now asserted in this litigation, this “underscores the conclusion that they are valid and patentable, or, at an absolute minimum, that Myriad is more likely than not to establish this factor at trial on the merits.” (Dkt. 5 at 14, n.5). The court disagrees that the claims the *AMP* plaintiffs chose to challenge in their suit constitute evidence of the validity of Plaintiffs’ other claims in this case.

(quoting United States’ Amicus Curiae Brief at 26). These circumstances “weigh[ed] against deferring to the PTO’s determination.” *Id.* On the same day the Supreme Court issued its *AMP* ruling, the USPTO delivered the following guidance to its own patent examiners:

As of today, naturally occurring nucleic acids are not patent eligible merely because they have been isolated. Examiners should now reject product claims drawn solely to naturally occurring nucleic acids or fragments thereof, whether isolated or not, as being ineligible subject matter under 35 U.S.C. § 101. Claims clearly limited to non-naturally-occurring nucleic acids, such as a cDNA or a nucleic acid in which the order of the naturally occurring nucleotides has been altered (e.g., a man-made variant sequence), remain eligible. Other claims, including method claims, that involve naturally occurring nucleic acids may give rise to eligibility issues and should be examined under the existing guidance in MPEP 2106, Patent Subject Matter Eligibility.³³

The court need not decide whether the Supreme Court’s refusal in *AMP* to defer to the USPTO’s past practice of awarding gene patents weakens the initial presumption of subject matter eligibility for Plaintiffs’ patent claims. Even if the presumption applies, the court concludes that Defendant has persuasively come forward with evidence to overcome it.

2. Section 101 Patent Eligible Subject Matter

The Supreme Court characterizes the subject matter eligibility inquiry as a “threshold test.” *Bilski*, 130 S. Ct. at 3225. Treating it as such in this case is particularly appropriate given the recent, closely related *AMP* litigation addressing the patent eligibility of several Myriad BRCA1 and BRCA2 patent claims. *Id.*; see also *Parker v. Flook*, 437 U.S. 584, 593 (U.S. 1978) (noting that “obligation to determine what type of discovery is sought to be patented must precede the determination of whether that discovery is, in fact, new or obvious.”).

Under 35 U.S.C. § 101, “[w]hoever invents or discovers any new and useful [1] process, [2] machine, [3] manufacture or [4] composition of matter, or any new and useful improvement

³³ <http://www.uspto.gov/patents/law/exam/myriad_20130613.pdf> (emphasis in original).

thereof, may obtain a patent therefor, subject to the conditions and requirements of this title.”

The first step in evaluating a patent’s subject matter eligibility is to “identify whether the claimed invention fits within one of the four statutory classes set out in § 101.” *Accenture Global Services, GmbH v. Guidewire Software, Inc.*, 728 F.3d 1336, 1341 (Fed. Cir. 2013) (citations omitted). The statute’s broad language suggests Congress “contemplated that the patent laws would be given wide scope.” *Chakrabarty*, 447 U.S. at 308.

But while claimed inventions may otherwise fall within § 101, important exceptions apply: “[l]aws of nature, natural phenomena, and abstract ideas’ are not patentable.” *Mayo*, 132 S. Ct. at 1293 (quoting *Diamond v. Diehr*, 450 U.S. 175, 185 (1981)) (other citations omitted). This is so irrespective of whether inventions are “[g]roundbreaking, innovative, or even brilliant,” *AMP*, 133 S. Ct. at 2117, or “just discovered,” because “they are the basic tools of scientific and technological work.” *Gottschalk v. Benson*, 409 U.S. 63, 67 (1972). Patents granted over these basic tools create “considerable danger” that their use would be “tie[d] up,” thereby “inhibit[ing] future innovations premised upon them.” *AMP*, 122 S. Ct. at 2116 (quoting *Mayo*, 132 S. Ct. at 1301). Such innovation-impeding monopolization is at odds with the “very point of patents, which exist to promote creation.” *AMP*, 133 S. Ct. at 2116 (citing *Chakrabarty*, 447 U.S. at 309).

This is illustrated in *Funk Bros. Seed Co. v. Kalo Inoculant Co.*, 333 U.S. 127 (1948), where the patentee discovered that certain strains of bacteria did not exert a mutually inhibitive effect on each other. The patentee created a mixed bacteria culture that could inoculate several species of leguminous plant seeds. This creation required ingenuity. Still, the Supreme Court concluded the discovery was patent ineligible because it was drawn to “the handiwork of nature,” where each of the combined bacteria possessed the same utility it had before it was

arguments advanced by Myriad in the *AMP* litigation, excludes from patent eligibility synthetic DNA that reflects naturally occurring BRCA1 and BRCA2 sequences. At every step, the *AMP* courts understood the isolated DNA at issue included both extracted genomic DNA and synthetic DNA. Thus, as described below, the court rejects Plaintiffs' argument that their Primer Claims are patent eligible because those claims relate to synthetic DNA.

But even if the court could conclude that the Supreme Court's *AMP* decision is not dispositive of the issue, the court independently reads the relevant authority—including the decisions in *AMP*, *Chakrabarty*, and *Funk Bros.* to compel the same conclusion, that Plaintiffs' Primer Claims are drawn to patent ineligible products of nature.

i. Under *AMP*, Synthetic DNA May be Patent Ineligible if it Reflects the Same Nucleotide Sequence as Naturally Occurring DNA

The *AMP* district court and the Federal Circuit panel understood that case to present squarely the issue of the patent eligibility of isolated DNA, including primers and probes. The *AMP* district court construed "isolated DNA" as Myriad urged, to "refer to a segment of DNA nucleotides existing separate from other cellular components normally associated with native DNA, including proteins and other DNA sequences comprising the remainder of the genome, and includes both DNA originating from the cell as well as DNA synthesized through chemical or heterologous biological means." 702 F. Supp. 2d at 217; '282 Patent col.19 ll.8-18; and '492 Patent col.17 ll.62—col.18 l.5. Judge Sweet noted that Myriad argued "native and isolated DNA" had "markedly different" functions because "isolated DNA may be used in applications for which native DNA is unsuitable, namely, in molecular diagnostic tests (e.g., as *probes*, *primers*, templates for sequencing reactions)" *Id.* at 230 (citations to Myriad's briefing omitted) (emphasis added). Judge Sweet nevertheless found all isolated DNA patent ineligible, as it is not "markedly different" from native, genomic DNA.

Likewise, the three Federal Circuit Judges on the *AMP* panel ruled on the patent eligibility of “isolated DNA,” which they understood to include primers and probes. As it had below, Myriad again argued to the Federal Circuit that “isolated DNAs, unlike native DNAs, can be used as primers and probes for diagnosing cancer.” 689 F.3d at 1325. Judge Lourie explained that “isolated DNA” is DNA that has been “cleaved . . . or synthesized to consist of just a fraction of a naturally occurring DNA molecule.” *Id.* at 1328. Judge Moore observed that the “smaller isolated DNA sequences” have “a variety of applications and uses in isolation that are new and distinct as compared to the sequence as it is in nature . . . [T]hese sequences can be used as *primers* . . . [*and*] *probes* . . .” *Id.* at 1341 (emphasis added). Judge Bryson did not explicitly use the term “primer” in relation to the non-cDNA isolated DNA. But his stated rationale for finding non-cDNA isolated DNA patent ineligible applies equally to DNA synthetically created for use as a primer as well as to extracted, genomic DNA. He noted that while isolated DNA molecules have been cleaved and have “terminal groups that differ from those found on naturally occurring genes,” they have the same sequence, code for the same proteins, and represent the same units of heredity. *Id.* at 1352. Their function is “dictated by the nucleotide sequence of the gene—a sequence that is determined by nature and that appears in nature exactly as it appears in the claimed isolated DNA.” *Id.*

Like Judge Bryson, the Supreme Court in *AMP* does not explicitly use the terms “primer” or “probe” when discussing the non-cDNA claims before it. Instead, the Court held that “a naturally occurring DNA segment is a product of nature and not patent eligible merely because it has been isolated, but that cDNA is patent eligible because it is not naturally occurring.” 133 S. Ct. at 2111. In conclusion, the Court stated: “We merely hold that genes and the information

they encode are not patent eligible under § 101 simply because they have been isolated from the surrounding genetic material.” *Id.* at 2120.

In contrast to Myriad’s (successfully) urged definition of “isolated DNA” throughout the *AMP* litigation, Plaintiffs now reverse course and argue that the isolated DNA the *AMP* Court found patent ineligible was only genomic DNA that is extracted from its natural environment, not synthetic DNA such as primers and probes. Pls.’ Reply Br. at 40 (the Court “used the term ‘isolation’ only to mean ‘extraction’ of genomic DNA.”). Springing from this interpretation, Plaintiffs suggest that if the Court found patent ineligible only isolated—now viewed by Plaintiffs to mean only genomic, extracted DNA—then the Court must have in a blanket fashion “affirmed the patent eligibility of synthetic DNA,” finding that “unlike isolated human genes, synthetic DNA is man-made and not a product of nature.” Pls.’ Mot. for Prel. Inj. at 4. Thus, Plaintiffs argue that “the very framework that the Court used—the distinction between naturally occurring DNA (which it held unpatentable) and artificially created, synthetic DNA, along with the methods of applying knowledge about the genes—leads to the conclusion that the claims Myriad asserts here, all of which fall into the latter category, are valid and enforceable.” (*Id.* at 14.)

Plaintiffs further contend the Federal Circuit utilized Plaintiffs’ now-urged ‘isolated DNA as distinct from synthetic DNA’ analysis. Plaintiffs claim that when the Federal Circuit heard the *AMP* case on remand, it held “that both isolated DNA and synthetic DNA were patent eligible under § 101, noting that ‘each of the claimed molecules represents a non-naturally occurring composition of matter.’” Pls.’ Mot. for Prel. Inj. at 7, n.1 (quoting Federal Circuit’s *AMP* decision, 689 F.3d at 1309).

133 S. Ct. at 2119.³⁵ If cDNA—which is clearly synthetic—is sometimes patent ineligible, then implicit in the Supreme Court’s decision is the conclusion that not all synthetic DNA is patent eligible.

Second, if the dispositive issue for patent eligibility was simply whether a DNA composition is synthetic, the Court’s analysis of cDNA might have begun and ended with the fact that cDNA is created in a laboratory. But the Court went well beyond this in its analysis, discussing its view of cDNA’s important uniqueness: that intervening introns are removed from the contiguous sequence in creating cDNA, and thus the “lab technician creates something new when cDNA is made”—something “distinct from the DNA from which it was derived.” *Id.* at 2119 (emphasis added).³⁶ The *AMP* Court was not focused simply on cDNA’s origin in a laboratory—isolated genomic DNA is extracted and purified in a laboratory as well. Rather, the Court focused on the fact that the cDNA’s contiguous sequence was altered in comparison to the sequence from which it was derived. The *AMP* Court later noted that its holding did not extend to “the patentability of DNA in which the order of the naturally occurring nucleotides has been altered.” 133 S. Ct. at 2120. The Court’s discussion of cDNA and altered DNA suggest that the Court employed, though not explicitly, something akin to the ‘magic microscope test’ the government urged before the Federal Circuit. *See* discussion of Federal Circuit *AMP* opinions *supra* Part I.E.2.

³⁵ This conclusion that cDNA may be patent ineligible echoes Judge Bryson’s opinion in which he concluded *Myriad* was not entitled to the broad patent protection for cDNA that it claimed. *See* 689 F.3d at 1356-57.

³⁶ Similarly, in *Chakrabarty*, the Court’s reasoning could have ended after noting that the bacterium at issue was “human-made” and therefore “synthetic.” 447 U.S. at 305. Instead, the Court went on to determine that the new bacterium might additionally have “markedly different characteristics from any [bacterium] found in nature.” *Id.* at 311.

Accordingly, this court’s best reading of *AMP* is that the Court concluded cDNA sometimes can be sufficiently different from naturally occurring matter as to merit patent eligibility. But non-cDNA isolated DNA is patent ineligible insofar as “the location and order of the nucleotides existed in nature”—whether that isolated DNA is cleaved, genomic DNA, or synthetic primers and probes with the same encoded information. Although the Supreme Court decision is not as explicit as the *AMP* lower court rulings in stating that non-cDNA isolated DNA includes primers and probes, this court reads the Supreme Court’s decision to harmonize with this proposition. 133 S. Ct. at 2116.

For these reasons, this court concludes that Plaintiffs are incorrect in contending that the *AMP* Court found all synthetic DNA to be patent eligible. Rather, this court interprets *AMP* to stand for the proposition that even synthetic, non-cDNA, isolated DNA is patent ineligible where it reflects the same nucleotide sequence as the genomic DNA.

ii. Under the Court's Independent Reading of *AMP*, *Funk Bros.*, and *Chakrabarty*, the Primer Claims are Drawn to Patent Ineligible Subject Matter Not Markedly Different from Naturally Occurring DNA

Even if the court could conclude that the *AMP* decision did not necessarily resolve the patent ineligibility of BRCA primers and probes, a narrow reading of the Supreme Court's decision and the cases cited therein, including *Funk Bros.*, and *Chakrabarty*, independently leads the court to conclude that Defendant has raised a substantial question concerning whether Plaintiffs' Primer Claims are drawn to patent ineligible subject matter. The claimed subject matter, although synthetically 'designed,' seems not "markedly different" from naturally occurring DNA.³⁷

Reading *AMP* narrowly, the Supreme Court clearly held that BRCA1 and BRCA2 genes and the information they encode are not patent eligible, even if isolated from other genetic material. 133 S. Ct. at 2120. This was so even though: 1) Myriad was first to uncover the location and sequence of the BRCA1 and BRCA2 genes, as these things "existed in nature before Myriad found them"; and 2) "isolating DNA from the human genome . . . creates a nonnaturally occurring molecule" due to the severing of chemical bonds, as Myriad's claims were expressed in terms of "the genetic information encoded in the BRCA1 and BRCA2 genes." *Id.* at 2118

³⁷ The court agrees with the observation of Defendant's expert Dr. David Pribnow that "the way Plaintiffs use 'design' implies that a scientist creates a BRCA primer sequence in a vacuum (or 'from scratch'). This is not accurate, scientifically. Primers are designed in reference to the natural[ly]-occurring sequence that is desired to be replicated following Watson-Crick base pairing." Pribnow 2nd Decl. at ¶ 6.

(also noting that if patents depended on creation of a unique molecule, a potential infringer might avoid patent claims on entire genes by “isolating a DNA sequence that included both the BRCA1 and BRCA2 gene and one additional nucleotide pair.”).

The Supreme Court’s conclusion in *AMP* flows from its prior jurisprudence in *Funk Bros.* and *Chakrabarty*. As discussed above, in *Funk Bros.*, an inventor discovered a mixed bacteria culture capable of inoculating several species of leguminous plants. The discovery was held patent ineligible because each bacterium in the culture retained its same natural structure and function, although in a new mixture. 333 U.S. at 131. In contrast, in *Chakrabarty*, a newly engineered bacterium capable of breaking down crude oil to help in oil spill cleanups was held patent eligible. 447 U.S. at 305-10. There, the inventor had not simply uncovered “hitherto unknown natural phenomenon,” but had produced a new composition “with markedly different characteristics from any found in nature and one having the potential for significant utility.” 447 U.S. at 310. Judge Bryson stated that *Chakrabarty* requires courts to “focus on two things” in evaluating patent eligibility: “(1) the similarity in structure between what is claimed and what is found in nature; and (2) the similarity in utility between what is claimed and what is found in nature.” 689 F.3d at 1354.

In this case, Plaintiffs’ Primers Claims are directed to compositions structurally similar to the DNA found in nature. They are drawn to pairs of single stranded primers with sequences exclusively derived or isolated from naturally occurring BRCA sequences located on chromosomes 17q (BRCA1) and 13 (BRCA2). Were they not, they could not bind to the target sequence found in a BRCA1 or BRCA2 gene. Stated more directly, the claimed primers must share a structural similarity with the naturally occurring DNA sequence if the primers are to serve the purpose claimed in Plaintiffs’ patents.

For example, claim 16 of the '282 Patent defines the structure of the claimed DNA primers with reference to the natural DNA sequence found in chromosome 17q (BRCA1). Claim 29 of the '492 Patent defines the structure of the claimed DNA primers by reference to the natural DNA sequence found in chromosome 13 (BRCA2). By definition, the claimed primers have the same nucleotide sequences as naturally occurring DNA.

The same can be said of claim 17 of the '282 Patent, which depends on claim 16, but “wherein said BRCA1 gene has the nucleotide sequence set forth in [exon only sequence];” as well as claim 30 of '492 Patent, which depends on claim 29, but where the “BRCA2 gene has the nucleotide sequence set forth in the [exon only] sequence.” These claims are to primers which are only 15 to 18 nucleotides and 25 to 30 nucleotides in length—typically much shorter than the naturally occurring exon-only regions in both BRCA1 and BRCA2 genes. For example, exon 11 of BRCA1 is about 3,400 nucleotides long, and an exon of BRCA2 is about 5,000 nucleotides long. Roa Decl. at ¶ 14. Thus, the claimed primers also have the same sequences as naturally occurring BRCA DNA.

This is true whether the nucleotide sequence is found in genomic DNA, or used in a primer or probe. Critically, it is this nucleotide sequence that gives DNA its unique role as an informational molecule. Pribnow Decl. at ¶¶ 19-22, 28; Kay Decl. at ¶ 13. The information set forth in a particular sequence of four nucleotides is the same whether the DNA is genomic or synthesized. Pribnow Decl. at ¶¶ 19-21, 27, 52-54; Pribnow 2nd Decl. at ¶¶ 5-11 (Dkt. 132). Like the isolated DNA at issue in *AMP*, the Primer Claims are drawn to compositions specifically expressed in terms of the nucleotide sequences derived or isolated from the naturally occurring BRCA1 and BRCA2 genes.

In addition to being structurally similar, the claimed primers are likewise similar in utility to naturally occurring DNA. As noted, the nucleotide sequences of the primers are necessarily derived or isolated exclusively from BRCA1 and BRCA2 sequences. This is so they will hybridize to complementary segments of the genes just as native DNA must, according to Watson-Crick pairing. In addition, during PCR, the primers function similarly to genomic DNA undergoing replication in the human body. Pribnow 2nd Decl. at ¶ 14. But for the similarity in utility between the primers and naturally occurring DNA sequences, the claimed primers would be incapable of serving the purposes necessary to Plaintiffs' patents.

Plaintiffs nevertheless contend their primers are “far removed” from anything appearing in nature, and thus are patent eligible. First, Plaintiffs argue that the Primer Claims are drawn to primer pairs designed to work in conjunction with one another. Plaintiffs allege that a pair of single stranded primers “is even further removed from being a product of nature” than a single stranded primer, as it is unlikely that a pair of primers would exist in nature that could hybridize to the same strand of DNA. Pls.’ Reply Br. at 47, n.16.

This argument is reminiscent of those made by the patent holder in *Funk Bros.* 333 U.S. at 131. The claimed invention there was a culture of naturally occurring bacterium which, when mixed, had beneficial effects. Still, the combination of the two did not transform the mixture into patent eligible subject matter because each type of bacteria in the culture simply “perform[ed] in their natural way,” serving the “ends nature originally provided. . . .” *Id.* at 131. Likewise, while the primers may be useful working in pairs, the fact that Myriad claims two of them together does not alter the fact that they remain patent ineligible products of nature, with each primer identical to the BRCA1 or BRCA2 nucleotide sequence from which it is derived, and carrying identical genetic information.

Second, Plaintiffs emphasize that their claimed primers are distinct from DNA in nature and thus patent eligible, because they are much shorter—between 15 to 18 or 25 to 30 nucleotides—than a naturally occurring BRCA1 or BRCA2 gene. Pls.’ Reply Br. at 47. But this argument is undermined by the fact that the *AMP* Court found claim 5 of Myriad’s ’282 Patent—directed to “an isolated DNA having at least 15 nucleotides of the DNA of claim 1”—was drawn to ineligible subject matter. That a segment of isolated DNA might be as short as 15 nucleotides was of no moment for the Court, which concluded that the isolated DNA in claim 5 was not patent eligible. *AMP*, 133 S. Ct. at 2113 (noting that “Myriad’s patents would, if valid, give it the exclusive right to isolate an individual's BRCA1 and BRCA2 genes (or any strand of 15 or more nucleotides within the genes) by breaking the covalent bonds that connect the DNA to the rest of the individual's genome.”).

Third, Plaintiffs appear to argue that the “necessarily incidental” chemical changes segments of genomic DNA undergo when extracted play a role in the patent eligibility analysis. Plaintiffs argue that their primers are different than naturally occurring DNA in that they “are limited according to the chemical properties designed and built . . . by the scientist . . . to prime a chemical reaction. . . .” Pls.’ Reply Br. at 48. Plaintiffs assert that because the chemical changes are ‘designed’ rather than ‘incidental to extraction,’ the primers are distinct from naturally occurring DNA. But Plaintiffs fail to make clear why this distinction matters in a § 101 analysis. Both extraction of genomic DNA and primer creation result in DNA that is not markedly different from naturally occurring DNA in either structure or function.

To the extent Plaintiffs urge their primers are distinct from naturally occurring DNA by virtue of ‘tags’ or terminating sequences placed at the ends of the primers and probes which are not derived from the BRCA sequences, this also fails to save Plaintiffs’ Primer Claims.

Specifically, the *AMP* Court noted the act of extracting DNA—severing its chemical bonds—“creates a nonnaturally occurring molecule,” but that this could not render patent eligible Myriad’s non-cDNA isolated DNA. *Id.* at 2118. Like the primers at issue here, the isolated DNA compositions Myriad claimed in *AMP* were “nonnaturally occurring,” but were drawn to natural genes and their sequences—“the information they encode.” *Id.* at 2120.

Fourth, Plaintiffs claim their primers are functionally different than genomic DNA because the claimed primer pairs can be used in PCR, or to find large deletions or duplications in a gene sequence. Pls.’ Reply Br. at 49. In essence, Plaintiffs argue that because primers can be used as primers, they have a utility beyond naturally occurring DNA. This ignores the fundamental reason primers are useful in PCR—because they function like natural DNA during replication, pairing predictably according to Watson-Crick principles. *See* Pribnow 2nd Decl. at ¶ 11 (primers’ “utility depends on the fact that a DNA segment used as a primer is structurally and functionally the same as a ‘native’ genomic DNA segment of the same sequence and length.”). That PCR with primers exploits this natural DNA function to a useful end does not render the function itself markedly different from that of naturally occurring DNA.

Relatedly, Plaintiffs contend that their Primer Claims should be patent eligible because cDNA, which can be patent eligible, “cannot be used as a primer.” Pls.’ Reply Br. at 49. The court does not find this argument persuasive. If the claimed primers share with naturally occurring DNA sufficient similarity in structure and utility, then it is immaterial whether another type of DNA is incapable of functioning as a primer.

For the foregoing reasons, the court concludes that Defendant has raised a substantial question concerning whether the Primer Claims are drawn to patent ineligible subject matter.

b. Section 101 Subject Matter Eligibility of the Method Claims

The court next turns to Plaintiffs' Method Claims, which are drawn to processes of comparing and analyzing BRCA1 and BRCA2 DNA. Plaintiffs contend that Defendant's testing infringes claims 7 and 8 of the '441 Patent, claim 4 of the '857 Patent, claim 5 of the '721 Patent, and claims 2 and 4 of the '155 Patent.

The court evaluates Plaintiffs' Method Claims in light of *Mayo* and the Federal Circuit's second *AMP* ruling, in which that court found all but one of the method claims at issue patent ineligible. Myriad chose not to cross-appeal that portion of the Federal Circuit's decision. While the Supreme Court's *AMP* opinion offers some guidance, the Court ruled only on Myriad's composition claims.³⁸

Thus, the court's analysis begins by reviewing the *AMP* Federal Circuit's ruling on the six Myriad method claims at issue there: claim 1 of the '999 Patent, claim 1 of the '001 Patent, claim 1 of '441 Patent, claims 1 and 2 of the '857 Patent, and claim 20 of the '282 Patent. The Federal Circuit concluded that five claims were patent ineligible as drawn to abstract, mental

³⁸ Concerning methods and applications, the *AMP* Court noted that if Myriad had "created an innovative method of manipulating genes while searching for the BRCA1 and BRCA2 genes, it could possibly have sought a method patent"; but the processes used by Myriad to isolate DNA at the time of Myriad's patents "were well understood, widely used, and fairly uniform insofar as any scientist engaged in the search for a gene would likely have utilized a similar approach" 133 S. Ct. at 2119-20 (quoting District Court's *AMP decision*, 702 F. Supp. 2d at 202-203). Plaintiffs imply that the *AMP* Court approved of Myriad's method claims, even though they were not before it. Plaintiffs argue that "as the Supreme Court implicitly recognized, the Method Claims asserted here by Myriad are distinguishable from the [patent ineligible] claims at issue in [*Mayo*], decided just prior to the [*AMP*] decision." Pls.' Mot. for Prelim. Inj. at 16, n.8. Plaintiffs also claim that the *AMP* Court "endorsed the validity of claims pertaining to synthetic DNA and methods of testing and using isolated genes in medical diagnosis and treatment." Pls.' Mot. for Prelim. Inj. at 7. The court does not read the Supreme Court's *AMP* ruling this way.

steps of “‘comparing’ and analyzing’ two gene [BRCA] sequences.” *Id.* at 1334.³⁹

These claims failed notwithstanding the fact that they were limited in application to comparing BRCA1 and BRCA2 sequences, which the Federal Circuit concluded were otherwise patent eligible. In so finding, the Federal Circuit noted that “the prohibition against patenting abstract ideas cannot be circumvented by attempting to limit the use of the formula to a particular technological environment.” 689 F.3d at 1334 (quoting *Bilski*, 130 S. Ct. at 3230 (other citations omitted)). Further, the court rejected Myriad’s attempt to read “into its method claims additional, allegedly transformative steps” not included in the claims, such as extracting and sequencing the BRCA DNA. 689 F.3d at 1335.

Myriad also argued in the Federal Circuit that the patent specifications suggested that the term “sequence” refers not just to information, but to a “physical DNA molecule, whose sequence must be determined before it can be compared.” *Id.* at 1334. The court disagreed, stating that while that “may be true,” the claims at issue “only recite mental steps, not the structure of physical DNA molecules.” *Id.* In fact, the Federal Circuit found the method claims before it less deserving of patent eligibility than the ineligible *Mayo* claims, where Myriad’s claims lacked even a “putatively transformative” determinative step—such as isolation and

³⁹ Only claim 20 of the ’282 Patent survived. It recites a method with steps of growing host cells transformed with an altered BRCA1 gene in the presence or absence of a possible cancer therapy, determining the ensuing growth rates of cells, and comparing the resulting growth rates. 689 F.3d at 1336. The claim was found patent eligible where it was drawn to a process for using transformed, non-naturally occurring cells which included “a foreign gene” and enhanced utility. *Id.* That the claim also includes the (presumably patent ineligible) “steps of determining the cells’ growth rates and comparing growth rates,” did not change the critical fact that the claim is based on a “man-made, non-naturally occurring transformed cell—patent eligible subject matter.” *Id.*

sequencing. All the court could discern were the mental processes of comparison and analysis.

Id. at 1335.

In this case, Plaintiffs agree that their asserted Method Claims “are all expressly limited to application of Plaintiffs’ discoveries of the sequence of the BRCA1 and BRCA2 genes,” and that they “generally recite methods for analyzing and/or comparing a patient’s BRCA1 or BRCA2 gene sequence to a normal reference or ‘wild-type’ sequence to determine if there are variations in the gene sequence.” Pls.’ Prop. Find. of Fact and Concl. of Law at 69 (Dkt. 152). Despite the striking initial similarities to the five patent ineligible method claims in *AMP*, Plaintiffs argue that the Method Claims here are patent eligible because they “employ specific laboratory testing processes that apply Myriad’s discovery of the BRCA1 and BRCA2 genes to develop physical steps that were not well-understood, routine, or conventional at the time the patents were filed.” Pls.’ Reply Br. at 54.⁴⁰ The Supreme Court’s *Mayo* decision is helpful to this court’s consideration of these arguments.

In *Mayo*, a unanimous Supreme Court held that a process focusing on a law of nature, natural phenomenon, or an abstract idea may be patent eligible, but only if it incorporates another “inventive concept”—“other elements or a combination of elements” sufficient to “ensure that the patent in practice amounts to significantly more than a patent upon the natural law itself.” *Id.* at 1294. The *Mayo* patents claimed processes for doctors to more effectively treat patients with autoimmune disorders using thiopurine drugs. Although scientists had previously “understood that the levels in a patient’s blood of certain metabolites . . . were

⁴⁰ Myriad also argues that the Method Claims are patent eligible because they employ “primers that are themselves patent eligible . . .” Pls.’ Reply Br. at 54. Because the court has held otherwise with regard to primers—an analysis that applies with equal force to probes—this contention fails.

computer which used the equation to continuously re-calculate the time to open the mold, and configuring the computer to signal a device to open the press at the right time. *Id.* at 177-79.

The *Mayo* Court noted that in *Diehr*, there had been no indication that “all these steps, or at least the combination of those steps, were in context obvious, already in use, or purely conventional.” 132 S. Ct. at 1299; *see also Bilski*, 130 S. Ct. at 3230 (noting *Diehr* process, viewed as a whole, was a “previously unknown method for molding raw, uncured synthetic rubber into cured precision products, using a mathematical equation to complete some of its several steps by way of a computer.”). Thus, the patentees were not seeking “to pre-empt the use of [the] equation,” but sought “only to foreclose from others the use of that equation in conjunction with all of the other steps in their claimed process.” *Id.* (quoting *Diehr*, 450 U.S. at 187). The other steps “apparently added to the formula something that in terms of patent law’s objectives had significance—they transformed the process into an inventive application of the formula.” *Id.*

In contrast, the Court in *Flook* held patent ineligible a method for adjusting ‘alarm limits’ in the catalytic conversion of hydrocarbons, which employed a mathematical algorithm to calculate the current alarm limits, then adjusted the system accordingly to reflect the new alarm-limit values. 437 U.S. at 585-587. The use of the patent ineligible algorithm—a law of nature—could not “support a patent unless there [was] some *other* inventive concept in its application.” *Id.* at 594 (emphasis added). But, the “only novel feature” of the claimed method was the algorithm. *Id.* at 588. Aside from it, the chemical processes, practice of monitoring the process variables, use of alarm limits, understanding that alarm limit values must be recomputed and readjusted, and the use of computers for automatic monitoring were all “well known.” *Id.* at 594. The Court held that appending this “conventional or obvious. . . post-solution activity” would not

“transform” the patent ineligible algorithm “into a patentable process.” *Id.* at 589. To allow this “exalts form over substance,” where a “competent draftsman could attach some form of post-solution activity to almost any mathematical formula” *Id.* at 590.

Following these cases, the court here must analyze: 1) whether the Method Claims at issue set forth an “inventive step” aside from the patent ineligible subject matter, and beyond “well-understood, routine, conventional activity previously engaged in” by those in the field; and 2) whether allowing the Method Claims risks preempting the use of a natural law, natural phenomenon, or abstract idea. *Mayo*, 123. S. Ct. at 1294, 1299.

i. The Inventive Concepts in the Method Claims are the BRCA1 and BRCA2 Sequences, and the Method Claims Otherwise Set Forth Well-Understood, Routine and Conventional Activity Engaged in by Scientists at the Time of Myriad’s Patent Applications

A close reading of Plaintiffs’ briefing reveals that the only “inventive concepts” in their Method Claims are the patent ineligible naturally occurring BRCA1 and BRCA2 sequences themselves. The claims contain no otherwise new process for designing or using probes, primers, or arrays beyond the use of BRCA1 or BRCA2 sequences in those processes.

Plaintiffs argue their claims “require the use of inventive DNA synthesized in a laboratory based upon knowledge about the BRCA1 and BRCA2 genes (e.g., gene-specific probes, primers, and arrays) . . . or pertain to such synthetic DNA compositions themselves. . . .” Pls.’ Mot. for Prel. Inj. at 16. Plaintiffs contend that the probes, primers, and arrays used in their Method Claims’ processes are inventive because they utilize the BRCA1 and BRCA2 sequences. Plaintiffs submit that because their claims “are limited to *specific application* of [amplifying, sequencing, probing, and screening] to the *new biomarkers* Myriad discovered,” those techniques could not have been previously “well-understood, routine, or conventional activity.” Pls.’ Reply Br. at 56-57 (emphasis in original). Where “the genes’ sequence was unknown,” it

was essentially impossible for anyone to “design a process to amplify segments of that unknown sequence . . . discovery of the BRCA1 and BRCA2 sequences was necessary to allow the creation of new primers, probes, and amplicons specifically designed for analysis of the new biomarkers.” Pls.’ Reply Br. at 57.⁴¹ Plaintiffs further argue that “[t]he steps in every claim thus require application of the previously-unknown BRCA1 or BRCA2 gene sequences to design primers or probes based on those specific sequences . . . the claims recite specific chemical assays that were not *routine* at the time the patents were filed because it was *impossible* as a practical matter to create or perform the assays without the knowledge of the BRCA1 and BRCA2 sequences.” Pls.’ Reply Br. at 58.

Aside from the patent ineligible, naturally occurring nucleotide sequence of the BRCA1 and BRCA 2 genes, the other steps set forth in the Method Claims are conventional activities that were well-understood and uniformly employed by those working with DNA at the time Myriad applied for its patents: DNA amplification, sequencing, comparisons, detecting alterations in sequences, and hybridizing probes to alleles. Tait Decl. at ¶ 37. The laboratory materials, reagents, and protocols to accomplish these activities were well known and widely available in the art by the time the first August 1994 patent application corresponding to the asserted patents

⁴¹ At times, Plaintiffs seem to claim that they developed new techniques, aside from first discovering the sequence of the BRCA1 and BRCA2 genes: “Myriad discovered a new biomarker, created new reagents and techniques that could now analyze this new biomarker, and *invented new methods of determining a patient’s risk of breast and ovarian cancer using these reagents and techniques.*” Pls.’ Mot. for Prel. Inj. at 16, n.6 (emphasis added). The court reads this passage as Plaintiffs merely asserting that the ‘newly invented methods’ were those previously understood and engaged in by those studying genes and hereditary disease, but with the new knowledge of the BRCA1 and BRCA2 genes. Setting aside the use of BRCA1 or BRCA2 sequences, Plaintiffs have not otherwise directed the court to any new methods or applications they invented for amplifying, sequencing, comparing, or hybridizing DNA generally.

had been filed. *Id.* at ¶ 31. Any scientist engaged in obtaining the sequence of a gene in a patient sample would rely on these techniques.

Plaintiffs' own patents acknowledge as much. For example, the '282 Patent provides that "the practice of the present invention employs, unless otherwise indicated, conventional techniques of chemistry, molecular biology, microbiology, recombinant DNA, genetics, and immunology." '282 Patent col.25 ll.50-55; Tait 2nd Decl. at ¶ 9; *see also* '441 Patent col.17 ll.20-27 ("These methods are well known and widely practiced in the art.").

At bottom, Plaintiffs ask the court to find that obtaining knowledge of the naturally occurring BRCA1 and BRCA2 sequences is somehow an inventive step sufficient to render the Method Claims patent eligible. This cannot be. The Federal Circuit already found that limiting the comparison to a specific technological field, "to just the BRCA genes or...to just the identification of particular alterations," still "fails to render the claimed process patent-eligible." 689 F.3d at 1334. Likewise, the Supreme Court has held that the "prohibition against patenting abstract ideas 'cannot be circumvented by attempting to limit the use of the formula to a particular technological environment.'" *Bilski*, 130 S. Ct. at 3230 (quoting *Diehr*, 450 U.S. at 191-92) (other citations omitted).

ii. The Method Claims Preempt Use of Laws of Nature

Moreover, *Mayo* cautions against finding patent eligible claims drawn to laws of nature and abstract ideas, particularly where allowance of the claims risks "tying up the use of the underlying natural laws, inhibiting their use in the making of further discoveries." 132 S. Ct. 1294. Here, if allowed, Plaintiffs' Method Claims would essentially foreclose the most widely used means to study and test for BRCA1 and BRCA2 genes.

To study a gene, geneticists generally must amplify a given DNA sample. Kay Decl. at ¶ 31. The most widely used means to amplify DNA is through PCR, which requires primers. Kay Decl. at ¶ 32; Pribnow Decl. at ¶ 70. The PCR process was patented in 1987, and since that time it has been critically important to DNA testing. Pribnow Decl. at ¶ 74. Probes, like primers, are short segments of DNA capable of hybridizing to DNA segments according to Watson-Crick pairing. Pribnow Decl. at ¶ 85. PCR using primers and probe hybridization are the means needed to determine and compare BRCA1 and BRCA2 sequences, and to conduct BRCA1 and BRCA2 tests. Tait Decl. at ¶¶ 48-51.

The recent decision in *Ariosa Diagnostic, Inc. v. Sequenom, Inc.* is instructive. 2013 WL 5863022 (N.D. Cal. Oct. 30, 2013).⁴² There, the patentees discovered that cell-free fetal DNA (cffDNA) is detectable in a pregnant woman's plasma or serum (plasma without platelets). *Id.* at *1. This was an important discovery. It provided a new method for prenatal diagnoses much less invasive and risky than previously used techniques, and more reliable than analyzing blood cell DNA. *Id.* The patent claims at issue were drawn to methods for detecting, amplifying, and testing paternally inherited nucleic acid (DNA and RNA). *Id.* at *2.

At the summary judgment stage, the parties agreed that neither cffDNA itself, nor its discovery in maternal plasma or serum, was patent eligible “because the presence of cffDNA in

⁴² This case is a continuation of *Aria Diagnostics, Inc. v. Sequenom, Inc.*, 726 F.3d 1296 (Fed. Cir. 2013). In *Aria Diagnostics*, the Federal Circuit vacated the district court's denial of Sequenom's motion for a preliminary injunction against Aria, now known as Ariosa. The district court based its denial on findings that Aria had raised substantial questions of both noninfringement and patent eligibility under § 101 because it found that Sequenom's claimed harm was not irreparable, and because the balance of harms and public interest favored Aria. The Federal Circuit faulted the district court's initial claim construction and infringement analysis. In remanding the case, the Federal Circuit instructed the trial court to consider patent eligibility in light of the recently-issued Supreme Court decision in *AMP*. *Id.* at 1304. Rather than revisit the preliminary injunction following the Federal Circuit's decision, the district court ruled on the parties' cross motions for summary judgment.

maternal plasma or serum is a natural phenomenon.” *Id.* at *8. Sequenom also acknowledged that its claims simply “apply ‘conventional techniques’ to the newly discovered natural phenomenon of cffDNA.” *Id.* Nevertheless, it argued the claimed methods were patent eligible because: 1) “they are novel *uses* of a natural phenomenon, rather than a patent on the natural phenomenon itself”; and 2) “the claims do not preempt all uses of cffDNA.” *Id.* at *7 (emphasis added).

The court disagreed, concluding that the amplifying and detection steps set forth in the claims were “well-understood, routine, and conventional activity” at the time of the discovery that cffDNA is present in maternal plasma or serum. That the steps were applied to patent ineligible cffDNA, could not imbue the claims with the requisite “‘inventive concept’ sufficient to ensure that the patent in practice amounts to significantly more than a patent upon the . . . natural phenomenon . . . itself.” *Id.* at *8 (quoting *Mayo*, 132 S. Ct. at 1294; and *Flook*, 437 U.S. at 594).

The *Sequenom* court first noted that “the only inventive part of the patent is that the conventional techniques of DNA detection known at the time of the invention are applied to paternally inherited cffDNA as opposed to other types of DNA.” *Id.* at *9. In other words, “the only inventive concept contained in the patent is the discovery of cffDNA, which is not patentable.” *Id.* The court cited *AMP* in its analysis, stating that while “the Supreme Court was not presented with method claims, the Court explained ‘[h]ad Myriad created an innovative method of manipulating genes while searching for the BRCA1 and BRCA2 genes, it could possibly have sought a method patent. But the processes used by Myriad to isolate DNA were well understood by geneticists at the time of Myriad’s patents. . . .’” *Id.* (citing *AMP*, 133 S. Ct. at 2119-20). In light of this, the *Sequenom* court observed that “looking at the claimed processes

as a whole, the only inventive component . . . is to apply those well-understood, routine processes to paternally inherited cffDNA, a natural phenomenon.” *Id.* at *10 (citing *Diehr*, 450 U.S. at 188 (noting claims must be considered as a whole)).

The court further concluded that the patent claims, if allowed, effectively preempted “all known methods of detecting cffDNA.” Although Sequenom presented evidence that other detection methods may have existed a few years after issuance of the patent, there was no evidence that the other methods were commercially viable:

If the alternative methods are not commercially viable, then the effect of the patent in practice would be to preempt all uses of the natural phenomenon. It is important to note that the ‘540 patent does not merely claim uses or applications of cffDNA, it claims methods for detecting the natural phenomenon. Because generally one must be able to find a natural phenomenon to use it and apply it, claims covering the only commercially viable way of detecting that phenomenon do carry a substantial risk of preempting all practical uses of it.

Id. at *11.

Sequenom’s parallels to this case are striking. There, the only “inventive concept” in the asserted claims was the discovery of a natural phenomenon in a particular location—paternal cffDNA in pregnant women’s plasma. The rest of the steps in the patent claims were routine, well understood activities in which scientists regularly engaged. If allowed, the claims risked precluding all practical uses of cffDNA.

Similarly, the inventive concept Plaintiffs here identify is the product of nature that they discovered the naturally occurring BRCA1 and BRCA2 sequences on chromosomes 17q and 13. Likewise, allowing Plaintiffs’ Method Claims here effectively preempts PCR involving BRCA1 or BRCA2 genes—the most widespread means of amplifying DNA to facilitate research and testing. Plaintiffs’ Method Claims effectively construct a wall around the naturally occurring

BRCA1 and BRCA2 genetic sequences, which, like the cffDNA in *Sequenom*, are naturally occurring, patent ineligible subject matter.

Plaintiffs attempt to distinguish *Sequenom* by pointing out that the cffDNA at issue in that case was known to exist before the patent holders discovered it in a new location. In contrast, Plaintiffs contend that “BRCA1 and BRCA2 genes were unknown before Plaintiffs’ inventions,” and that the Method Claims here involve not only “the use of previously unknown markers (BRCA1 and BRCA2 genes) but also require the use of previously unknown BRCA1 and BRCA2 specific primers and probes invented by Myriad.” Pls.’ Response to Supp. Auth. at 2 (Dkt. 164). Thus, Plaintiffs claim “numerous” inventive concepts: “BRCA1 and BRCA2 cDNA sequences, BRCA1- and BRCA2-specific primer pairs, and methods of using those specific primer pairs, as well as probes, to diagnose breast and ovarian cancer.” *Id.* at 3.

The court is not persuaded by Plaintiffs’ efforts to distinguish *Sequenom*. As noted above, the BRCA1 and BRCA2 sequences and the primers setting forth those sequences are patent ineligible products of nature. Plaintiffs cite no legal authority for their position that their own patent ineligible discovery—the naturally occurring BRCA1 and BRCA2 sequences—should benefit them in a § 101 analysis of their Method Claims because they were the first to find the previously unmapped sequences of the BRCA1 and BRCA2 genes on chromosomes 17q and 13.⁴³

Based upon the foregoing, the court concludes that Defendant has raised a substantial question concerning the Method Claims’ subject matter eligibility for patent. Accordingly,

⁴³ Plaintiffs also make this argument to distinguish the patent ineligible subject matter underlying the method claims at issue in *Mayo*, where Plaintiffs contend that the “patentee’s *only* contribution to the art was a refined therapeutic range.” Pls.’ Reply Br. at 56 (emphasis in original).

Plaintiffs have not established that they are likely to succeed on the merits of their claims. The court concludes that Plaintiffs are not entitled to a preliminary injunction at this time.

D. Balancing of Equities or Hardships and the Public Interest

Although the court has determined that Plaintiffs are not entitled to a preliminary injunction, the court addresses below the remaining equitable preliminary injunction factors: the balance of the parties' equities or hardships, and the public interest.

1. Balance of Hardships

The court must weigh the "harm that will occur to the [Plaintiffs] from the denial of the preliminary injunction with the harm that [Defendant] will incur if the injunction is granted." *Hybridtech, Inc. v. Abbott Labs.*, 849 F.2d 1446, 1457 (Fed. Cir. 1988) (citations omitted). As noted above, the court has already concluded that without an injunction, Plaintiffs are likely to suffer irreparable harm in the form price erosion, loss of market share, and loss of the remainder of their exclusive patent terms. Nevertheless, Defendant contends that Myriad's strong financial position will soften any hardship Plaintiffs might suffer without an injunction, while an injunction would cause Defendant to suffer much more acute harm.

First, Defendant points to the loss of its "head start" as the first testing company offering alternatives to Myriad BRCA tests. Hampton Decl. at ¶¶ 58-60. The court does not find this argument persuasive. That Myriad is a large company and can survive an injunction does not compel the court to conclude that Defendant's loss of a head start outweighs Plaintiffs' loss of return on its years of work and substantial investment commercializing BRCA testing.

Defendant further contends that if an injunction were to issue, it likely would be forced out of business. That an accused infringer may go out of business if an injunction issues is something the court may consider balancing the parties' hardships. *Aria Diagnostics*, 726 F.3d

at 1305 (consideration of an accused infringer's loss of business is a proper consideration in preliminary injunction determination) (citations omitted). In advance of its announcement that it would offer genetic testing, including BRCA1 and BRCA2, Defendant invested an estimated \$46.7 million in capital resources, expanding its laboratory and hiring an additional 110 employees. Hampton Decl. at ¶ 57; Chao Decl. at ¶¶ 71-73. With Myriad's recent launch of myRisk, a multi-gene panel test similar to Defendant's CancerNext test, an injunction would leave consumers with little reason to order one of Defendant's tests, not including BRCA testing, when they could order myRisk, which does. Hampton Decl. at ¶ 61.

But the destruction of an infringer's business does not necessarily outweigh the patentee's injury in lost investments in developing and marketing a patented product. *Robert Bosch*, 659 F.3d at 1156 (infringer cannot "escape a[] [permanent] injunction simply because it is smaller than the patentee or because its primary product is an infringing one.") (citations omitted)). The Federal Circuit has made clear that when an accused infringer undertakes a "calculated risk" to launch a potentially infringing product, notions of fairness are not offended if the potential infringer bears the risk of their "almost entirely preventable" harms. *Sanofi-Synthelabo v. Apotex, Inc.*, 470 F.3d 1368, 1383 (Fed. Cir. 2006) (citations omitted). Similarly, in the case Plaintiffs cite, *Ortho Pharm. Corp. v. Smith*, 15 U.S.P.Q.2d 1856, 1863 (E.D. Pa. 1990), the district court concluded that the balance of hardships favored the patentee where the hardship the accused infringer might suffer, lost investments "in developing and preparing to market an infringing product," was "attributable solely to Ortho's calculated decision to bring [its product] to market prematurely"

Although relevant, the above-cited cases in which accused infringers are criticized for making costly product investments arise out of infringement of valid or likely-valid patents.⁴⁴ In contrast, Defendant here has succeeded in raising a substantial question concerning the subject matter eligibility of Plaintiffs' asserted patent claims. And Defendant waited to launch BRCA testing until the years-long *AMP* litigation had concluded and had at least cast considerable doubt on the subject matter eligibility of Plaintiffs' patent claims covering BRCA primers, probes, and some methods for comparing and analyzing BRCA sequences. The court finds that Defendant appears to have acted with some caution in timing its BRCA testing launch after the conclusion of the *AMP* litigation. Further, the court has concluded that Defendant's belief in the appropriateness of its launch following the Supreme Court's *AMP* ruling was not misplaced.

In view of the foregoing, the court concludes that Defendant's potential hardship in losing its entire business outweighs the hardship Myriad may suffer in terms of price erosion, lost market share, and the loss of the remainder of its patents' exclusive terms, which begin to expire in the coming months. Although Plaintiffs will suffer economic harm without an injunction, Myriad has enjoyed an exclusive monopoly in the BRCA1 and BRCA2 testing market for nearly two decades, and its own financial forecasts show that it expects to see increased revenue growth this year. Even without an injunction, Plaintiffs will undoubtedly continue to benefit from Myriad's expertise, market strength, and brand name recognition.

⁴⁴ In *Robert Bosch*, the defendant was found to have infringed some of the plaintiff's valid patents, but the district court denied a motion for a permanent injunction on the grounds that the plaintiff had failed to show irreparable harm. 659 F.3d at 1145. The Federal Circuit reversed and remanded. In *Sanofi-Synthelabo*, the Federal Circuit affirmed the district court's conclusion that the patentee was likely to succeed on the merits of showing that the patent at issue was valid and infringed. 470 F.3d at 1374-81. In *Ortho*, the district court considered the patentee's entitlement to a preliminary injunction, and concluded that it had shown a likelihood of success on the merits—fending off the accused infringer's validity challenges. 15 U.S.P.Q.2d at 1863.

Notwithstanding the court's conclusion that Plaintiffs will suffer irreparable injury without an injunction, the court nevertheless concludes that the balance of hardships factor tips slightly in Defendant's favor, and provides further reason not to impose a preliminary injunction at this time.

2. Public Interest

This factor requires the court to focus on whether “there exists some critical public interest that would be injured by the grant of preliminary relief.” *Hybritech*, 849 F.2d at 1458 (citations omitted). Plaintiffs correctly submit that there generally exists a strong public interest in upholding a patentee’s exclusive rights. This public interest yields from recognition that “[t]he patent system represents a carefully crafted bargain that encourages both the creation and the public disclosure of new and useful advances in technology, in return for an exclusive monopoly for a limited period of time.” *Pfaff v. Wells Elecs., Inc.*, 525 U.S. 55, 63 (1998). But the public’s interest in preserving patent rights will not always trump other considerations, especially when public health issues are at stake. *See, Hybritech*, 849 F.2d at 1458 (in case involving patents relating to medical tests for variety of conditions affirming district court’s conclusion that public interest would be served by injunction on sales of certain tests, but not tests for hepatitis and cancer). *cf. Dippin’ Dots v. Mosey*, 44 U.S.P.Q.2d 1812, 1818-19 (N.D. Tex. 1997) (finding public interest not affected by injunction involving patents on “ice cream, not heart valves, medical catheters, drug therapies or the cure for the common cold.”).

In this case, both sides make compelling arguments that the public interest favors them. Plaintiffs persuasively argue that the public interest would be served by protecting Plaintiffs' exclusive patent rights, particularly where they have, while exercising these rights, invested over \$500 million to improve Myriad's testing products; develop an extensive database of variant

classifications; create a market wherein third-party payors will reimburse testing costs; and provide testing to over one million patients. Ford Decl. at ¶¶ 5-7, 14. Plaintiffs contend no critical public health interest exists necessitating Defendant's testing because Myriad's BRACAnalysis has with "unparalleled reliability" met the need for BRCA1 and BRCA2 testing. Pls.' Reply Br. at 128. In short, Plaintiffs contend their advancements and investments can be credited with saving numerous lives. And while there may have been delays in getting Myriad's large rearrangement analysis and multi-gene panel test products to market, Myriad now offers large rearrangement testing in its BART test, and has recently launched myRisk, a multi-gene panel test very similar to Defendant's Cancer First.⁴⁵

But Plaintiffs' testing is much more expensive than many of Defendant's offered tests, and in some cases requires separate billing for point mutation and large rearrangement analyses.⁴⁶ Here, the court is concerned that Myriad's BART large rearrangement testing, which offers follow-up reassurance concerning a negative or inconclusive BRACAnalysis point mutation test, is neither offered as a matter of course, nor covered by third-party payors for all patients. Some women who should obtain this testing must pay out-of-pocket for it, if they can afford it. Further, although the evidence before the court is in conflict, Defendant has set forth

⁴⁵ Defendant vigorously argues that Myriad delayed getting to market its BART (large rearrangement) testing, and that (at the time briefing was completed) Myriad did not offer multi-gene panel testing. Def's. Opp. Memo. at 98-101. For purposes of the court's analysis of whether there is a critical public health interest in allowing Defendant's testing, the court neither finds the alleged delay in getting BART to market nor Myriad's past failure to offer a multi-gene panel dispositive. Myriad now offers large rearrangement testing in BART and has begun offering a multi-gene panel test, myRisk, that is similar to Defendant's testing.

⁴⁶ Myriad's BRACAnalysis and BART tests are often billed separately, while Defendant's BRCA1 and BRCA2 testing automatically includes point mutations and large rearrangement analysis for one price. Chao Decl. at ¶ 17.

evidence suggesting that Myriad testing is “out of network” with Tri-Care, a health care program serving United States military service personnel and their families.

Notwithstanding this, Plaintiffs have identified several ways in which they mitigate price barriers for many patients. First, Myriad has secured and maintained in-network contracts with more than 530 private payors to ensure that more patients have insurance coverage for testing, and the lowest possible out-of-pocket expense. Ford 2nd Decl. at ¶ 4 (Dkt. 27). Second, Myriad has developed four patient assistance programs for individuals meeting clinical guidelines who cannot afford testing: (1) free testing for those patients who meet clinical criteria, are uninsured, and under a set income level; (2) “capping” of out-of-pocket costs to \$375 for qualifying low-income patients with insurance; (3) interest-free financing; and (4) donations and discounted testing for charitable organizations. Over the past five years, these programs have benefitted more than 35,000 women. *Id.* The court notes this is not an insignificant undertaking.

Still, Defendant compellingly argues that Plaintiffs’ exclusive patent rights have, in many ways, hindered rather than promoted innovation in this area of significant public health interest. Defendant’s expert Dr. Joseph Stiglitz observes that the sequencing of BRCA1 and BRCA2 was inevitable at the time Myriad first discovered the natural location of the genes, and the promise of patents provided an unnecessary incentive, given the hotly-contested “race” in which Myriad and others were engaged.⁴⁷ Stiglitz Decl. at ¶ 36 (Dkt. 53). Moreover, the practical result of Myriad’s patents has been to hinder or halt follow-up research, data sharing, patient testing, and the creation of additional and more affordable technologies for BRCA1 and BRCA2 testing. For example, since about 2005, Myriad has declined to publicly share critical information regarding

⁴⁷ Dr. Stiglitz also argues that gene patents are particularly pernicious in general because they amount to patents on basic scientific knowledge. The court views this argument as more relevant to a § 101 subject matter eligibility analysis.

its classification of variants, including with its own patients. Instead, Myriad retains that information in a private database. In so doing, Myriad distorts rather than serves the patent system's goal of public disclosure in exchange for exclusive rights. In this way, Myriad has chosen a commercial path that turns much of our patent system policy on its head.

In short, there are compelling public interest arguments advanced by both sides. The court concludes that neither side has shown that the public interest mandates either the imposition or denial of Plaintiffs' requested injunction.

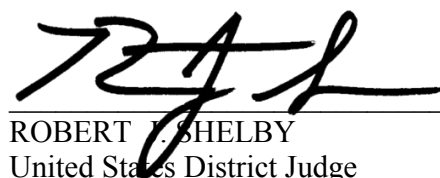
III. CONCLUSION

The court concludes Plaintiffs are not entitled to the extraordinary remedy of a preliminary injunction to halt Defendant from selling its own BRCA1 and BRCA2 genetic tests. Although Plaintiffs have shown that they are likely to suffer irreparable harm through erosion of their test pricing structure, loss of their share of the testing market, and loss of their exclusive patent terms if an injunction does not issue, Defendant has raised a substantial question concerning whether Plaintiffs' Primer and Method Claims are directed toward patent eligible products of nature and abstract ideas under 35 U.S.C. § 101. In light of Defendant's persuasive showing, Plaintiffs are unable to establish that they are likely to succeed on the merits of their claims. Nor have Plaintiffs demonstrated that the remaining equitable factors support issuance of the requested injunction.

For these reasons, Plaintiffs' Motion for Preliminary Injunction (Dkt. 5) is DENIED.

DATED this 10th day of March, 2014.

BY THE COURT:



ROBERT J. SHELBY
United States District Judge



US005654155A

United States Patent [19]**Murphy et al.**[11] **Patent Number:** **5,654,155**[45] **Date of Patent:** **Aug. 5, 1997**[54] **CONSENSUS SEQUENCE OF THE HUMAN BRCA1 GENE**[75] Inventors: **Patricia D. Murphy**, Slingerland, N.Y.;
Antonette C. Allen, Millersville, Md.;
Christopher P. Alvares, Potomac, Md.;
Brenda S. Critz, Frederick, Md.; **Sheri J. Olson**, Arlington, Va.; **Denise B. Schelter**, Burtonsville; **Bin Zeng**, Rockville, both of Md.[73] Assignee: **OncorMed, Inc.**, Gaithersburg, Md.[21] Appl. No.: **598,591**[22] Filed: **Feb. 12, 1996**[51] **Int. Cl.**⁶ **C12Q 1/68**; C12P 19/34;
C07H 21/04; C07H 21/02[52] **U.S. Cl.** **435/6**; 435/91.2; 536/23.1;
536/24.3; 536/24.33[58] **Field of Search** 435/6, 91.2; 536/23.1,
536/24.3, 24.33[56] **References Cited****U.S. PATENT DOCUMENTS**

4,458,066 7/1984 Carothers et al. 536/27

FOREIGN PATENT DOCUMENTS0699754A1 3/1996 European Pat. Off. .
705903A 4/1996 European Pat. Off. .
0705903A1 4/1996 European Pat. Off. .
0705902A1 4/1996 European Pat. Off. .**OTHER PUBLICATIONS**Miki et al. *Science* 266: 66-71 1994.
Friedman et al. *Nature Genetics* 8: 399-404 Dec. 1994.
Scriber et al. In *Variation in the Human Genome*, Ciba Foundation Symposium 197, John Wiley & Sons eds, New York, pp. 73-96. 1996.
Roswell, S., et al., *American Journal of Human Genetics* 55:861-865, (1994).
Miki, Y., et al., *Science* 266:66-71, (1994).
Friend, S., et al., *Nature Genetics* 11:238, (1995).
Beaudet, Arthur L., et al., A Suggested Nomenclature for Designating Mutations. *Human Mutation* 2:245-248, 1993.
Friedman, Lori S., et al., Confirmation of BRCA1 by Analysis of Germline Mutations Linked to Breast and Ovarian Cancer. *Nature Genetics* 8: 399-404, Dec. 1994.
Shattuck-Eidens, Donna, et al., A Collaborative Survey of 80 Mutations in the BRCA1 Breast and Ovarian Cancer Susceptibility Gene. *JAMA* 273:535-541, Feb. 15, 1995.
Ford, D., et al., The Genetics of Breast and Ovarian Cancer. *Genetics Heterogeneity* 5:805-811.Szabo, Csilla I., et al., *Inherited Breast and Ovarian Cancer. Human Molecular Genetics* 4:1811-1817, 1995.

Plummer, S.J., et al., Detection of BRAC1 Mutations by the Protein Truncation Test 4:1989-1991, 1995.

Chen, Y., et al., Aberrant Subcellular Localization of BRCA1 in Breast Cancer. *Science* 270: 789-791, Nov. 3, 1995.Gayther, Simon A., et al., Germline Mutations of the BRCA1 Gene in Breast and Ovarian Cancer Families Provide Evidence for A Genotype-phenotype Correlation. *Nature Genetics* 11: 428-433, Dec. 1995.

Durocher, F., et al., Comparison of BRCA1 Polymorphisms, Rare Sequence Variants and/or Missense Mutations in Unaffected and Breast/Ovarian Cancer Populations 5: 835-842, 1996.

Berman, D.B., et al., Two Distinct Origins of a Common BRCA1 Mutation in Breast-Ovarian Cancer Families: A Genetic Study of 15 185delAG-Mutation Kindreds. *Am. J. Hum. Genet.* 58:1166-1176, 1996.Schultz, D.C., et al., Identification of Two Candidate Tumor Suppressor Genes on Chromosome 17p13.3¹. *Cancer Research* 56:1997-2002, May 1, 1996.Couch, Fergus J., et al., Mutations and Polymorphisms in the Familial Early-Onset Breast Cancer (BRCA1) Gene. *Human Mutation* 8:8-18, 1996.Sanger, F., et al., *J. Mol. Biol.* 42:1617, (1980).Beaucage, et al., *Tetrahedron Letter*. 22:1859-1862.Maniatis, et al., In *Molecular Cloning: A Laboratory Manual*, Cold Spring, NY, pp. 280-281, (1982).Conner, et al., *Proc. Natl. Acad. Sci. U.S.A.* 80:278, (1983).Saiki, et al., *Bio/Technology* 3: 1088-1012 (1985).Landegren, et al., *Science* 241:1007, (1988).Landgren, et al., *Science* 242:229-237, (1988).McPherson, M.J., et al., *PCR. A Practical Approach*. ILR Press, Eds. (1992).Easton, et al., *American Journal of Human Genetics* 52: 678-701, (1993).*Primary Examiner*—W. Gary Jones*Assistant Examiner*—Dianne Rees*Attorney, Agent, or Firm*—R. Thomas Gallegos; Albert P. Halluin[57] **ABSTRACT**

A consensus DNA sequence has been determined for the BRCA1 gene. As has been seven polymorphic sites and their rates of occurrence in normal BRCA1 genes. The consensus gene BRCA1^(omi) and the seven polymorphic sites will provide greater accuracy and reliability for genetic testing. One skilled in the art will be better able to avoid misinterpretations of changes in the gene, determine the presence of a normal gene, and of mutations, and to classify tumors.

4 Claims, No Drawings**A000107**

5,654,155

1

CONSENSUS SEQUENCE OF THE HUMAN BRCA1 GENE

FIELD OF THE INVENTION

This invention relates to a gene which has been associated with breast and ovarian cancer where the gene is found to be mutated. More specifically, this invention relates to the most likely sequence (i.e. "Consensus Normal DNA sequence") for the BRCA1 gene (BRCA1^(omi) SEQ. ID. NO: 1) in normal individuals.

BACKGROUND OF THE INVENTION

It has been estimated that about 5–10% of breast cancer is inherited Rowell, S., et al., *American Journal of Human Genetics* 55:861–865 (1994). Located on chromosome 17, BRCA1 is the first gene identified to be conferring increased risk for breast and ovarian cancer. Miki et al., *Science* 266:66–71 (1994). Mutations in this "tumor suppressor" gene are thought to account for roughly 45% of inherited breast cancer and 80–90% of families with increased risk of early onset breast and ovarian cancer. Easton et al., *American Journal of Human Genetics* 52:678–701 (1993).

Locating one or more mutations in the BRCA1 region of chromosome 17 provides a promising approach to reducing the high incidence and mortality associated with breast and ovarian cancer through the early detection of women at high risk. These women, once identified, can be targeted for more aggressive prevention programs. Screening is carried out by a variety of methods which include karyotyping, probe binding and DNA sequencing.

In DNA sequencing technology, genomic DNA is extracted from whole blood and the coding regions of the BRCA1 gene are amplified. Each of the coding regions is sequenced completely and the results are compared to the normal DNA sequence of the gene.

The BRCA1 gene is divided into 24 separate exons. Exons 1 and 4 are noncoding, in that they are not part of the final functional BRCA1 protein product. The BRCA1 coding region spans roughly 5600 base pairs (bp). Each exon consists of 200–400 bp, except for exon 11 which contains about 3600 bp. To sequence the coding region of the BRCA1 gene, each exon is amplified separately and the resulting PCR products are sequenced in the forward and reverse directions. Because exon 11 is so large, we have divided it into twelve overlapping PCR fragments of roughly 350 bp each (segments "A" through "L" of BRCA1 exon 11).

Many mutations and normal polymorphisms have already been reported in the BRCA1 gene. A world wide web site has been built to facilitate the detection and characterization of alterations in breast cancer susceptibility genes. Such mutations in BRCA1 can be accessed through the Breast Cancer Information Core at: http://www.nchgr.nih.gov/dir/lab_transfer/bic. This data site became publicly available on Nov. 1, 1995. Friend, S. et al. *Nature Genetics* 11:238, (1995).

The genetics of Breast/Ovarian Cancer Syndrome is autosomal dominant with reduced penetrance. In simple terms, this means that the syndrome runs through families such that both sexes can be carriers (only women get the disease but men can pass it on), all generations will likely have breast/ovarian or both diseases and sometimes in the same individual, occasionally women carriers either die young before they have the time to manifest disease (and yet offspring get it) or they never develop breast or ovarian cancer and die of old age (the latter people are said to have

2

"reduced penetrance" because they never develop cancer). Pedigree analysis and genetic counseling is absolutely essential to the proper workup of a family prior to any lab work.

Until now, only a single normal sequence for BRCA1 has been available for comparison. That sequence is available as GenBank Accession Number U15595. There is a need in the art, therefore, to have available a normal sequence which represents most likely BRCA1 sequence to be found in the majority of the normal population, the (i.e. "Consensus Normal DNA sequence"). A Consensus Normal DNA sequence will make it possible for true mutations to be easily identified or differentiated from polymorphisms. Identification of mutations of the BRCA1 gene and protein would allow more widespread diagnostic screening for hereditary breast and ovarian cancer than is currently possible.

A consensus normal gene sequence of the BRCA1 is provided which more accurately reflects the most likely sequence to be found in a subject. Use of the consensus normal gene sequence (BRCA1^(omi)SEQ ID. NO: 1), rather than the previously published BRCA1 sequence, will reduce the likelihood of misinterpreting a "sequence variation" found in the normal population with a pathologic "mutation" (i.e. causes disease in the individual or puts the individual at a high risk of developing the disease). With large interest in breast cancer predisposition testing, misinterpretation is particularly worrisome. People who already have breast cancer are asking the clinical question: "is my disease caused by a heritable genetic mutation?" The relatives of the those with breast cancer are asking the question: "Am I also a carrier of the mutation my relative has? Thus, is my risk increased, and should I undergo a more aggressive surveillance program."

SUMMARY OF THE INVENTION

The present invention is based on the discovery of the most likely sequence to be found in normal human individuals for the BRCA1 gene.

It is an object of the invention to provide a consensus sequence for the normal BRCA1 gene, i.e. the consensus Sequence having the more commonly occurring nucleotides where normal polymorphisms occur.

It is another object of the invention to provide a consensus normal protein sequence of the BRCA1 protein

It is another object of the invention to provide a list of the codon pairs which occur at each of seven polymorphic points on the normal BRCA1 gene.

It is another object of the invention to provide the rates of occurrence for the codons.

It is another object of the invention to provide a method wherein BRCA1, or parts thereof, is amplified with one or more oligonucleotide primers.

It is another object of this invention to provide a method of avoiding misinterpretation of changes which a laboratory may find in the BRCA1 gene.

It is another object of this invention to provide a method of identifying individuals who carry no mutation(s) of the BRCA1 gene and are therefore at no increased risk or susceptibility to breast or ovarian cancer based on a finding that the individual does not carry an abnormal BRCA1 gene.

It is another object of this invention to provide a method of identifying a mutation leading to predisposition or higher susceptibility to breast or ovarian cancer.

It is another object of the invention to provide a method of classifying a tumor for diagnostic and prognostic purposes.

5,654,155

3

There is a need in the art for a consensus normal sequence of the BRCA1 gene and for the consensus normal protein sequence of BRCA1 as well as for an accurate list of codons which occur at polymorphic points on a normal sequence. A person skilled in the art of genetic susceptibility testing will find the present invention useful for:

- a) identifying individuals having a normal BRCA1 gene, who are therefore not at risk or have no increased susceptibility to breast or ovarian cancer from a BRCA1 mutation;
- b) avoiding misinterpretation of normal polymorphisms found in the normal BRCA1 gene;
- c) determining the presence of a previously unknown mutation in the BRCA1 gene.
- d) identifying a mutation which indicates a predisposition or higher susceptibility to breast or ovarian cancer; or for
- e) classifying a tumor for diagnostic and prognostic purposes.
- f) performing gene repair.

DETAILED DESCRIPTION OF THE INVENTION

Definitions

The following definitions are provided for the purpose of understanding this invention.

"Consensus Normal Sequence" refers to the nucleic acid or protein sequence, the nucleic or amino acids of which are known to occur with high frequency in a population of individuals who carry the gene which codes for a normally functioning protein, or which nucleic acid itself has normal function.

"Consensus normal DNA sequence" also called "consensus normal gene sequence" refers to a nucleic acid sequence, the nucleic acid of which are known to occur at their respective positions with high frequency in a population of individuals who carry the gene which codes for a normally functioning protein, or which itself has normal function.

"Consensus Normal Protein Sequence" refers to the protein sequence, the amino acids of which are known to occur with high frequency in a population of individuals who carry the gene which codes for a normally functioning protein. "BRCA1^(om)(SEQ ID NO: 1)" refers to a consensus sequence for the BRCA1 gene. The consensus sequence was found by end to end sequencing of the BRCA1 gene from 5 individuals randomly drawn from the population and found to have no family history of breast or ovarian cancer. The sequenced gene was found not to contain any mutations. The consensus was determined by calculating the frequency with which nucleic acids occur and inserting the nucleic acid with highest frequency of occurrence at each polymorphic site. In some cases the insertion of a nucleic acid at a polymorphic site indicated a codon change and a change of amino acid from previously published information. In other cases the frequency of occurrence of a nucleic acid was found to differ from the published frequency.

The term "primer" as used herein refers to a sequence comprising about 20 or more nucleotides of the BRCA1 gene.

A "target polynucleotide" refers to the nucleic acid sequence of interest e.g., the BRCA1 encoding polynucleotide. Other primers which can be used for primer hybridization will be known or readily ascertainable to those of skill in the art.

4

The term "substantially complementary to" or "substantially the sequence" refers to primer sequences which hybridize to the sequences provided under stringent conditions and/or sequences having sufficient homology with BRCA1 sequences, such that the allele specific oligonucleotide primers hybridize to the BRCA1 sequences to which they are complementary.

The term "isolated" as used herein includes oligonucleotides substantially free of other nucleic acids, proteins, lipids, carbohydrates or other materials with which they may be associated. Such association is typically either in cellular material or in a synthesis medium.

Sequencing

Any nucleic acid specimen, in purified or non-purified form, can be utilized as the starting nucleic acid or acids, providing it contains, or is suspected of containing, the specific nucleic acid sequence containing a polymorphic locus. Thus, the process may amplify, for example, DNA or RNA, including messenger RNA, wherein DNA or RNA may be single stranded or double stranded. In the event that RNA is to be used as a template, enzymes, and/or conditions optimal for reverse transcribing the template to DNA would be utilized. In addition, a DNA-RNA hybrid which contains one strand of each may be utilized. A mixture of nucleic acids may also be employed, or the nucleic acids produced in a previous amplification reaction herein, using the same or different primers may be so utilized. See TABLE II. The specific nucleic acid sequence to be amplified, i.e., the polymorphic locus, may be a fraction of a larger molecule or can be present initially as a discrete molecule, so that the specific sequence constitutes the entire nucleic acid. It is not necessary that the sequence to be amplified be present initially in a pure form; it may be a minor fraction of a complex mixture, such as contained in whole human DNA.

DNA utilized herein may be extracted from a body sample, such as blood, tissue material and the like by a variety of techniques such as that described by Maniatis, et. al. in *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor, N.Y., p 280-281, 1982). If the extracted sample is impure, it may be treated before amplification with an amount of a reagent effective to open the cells, or animal cell membranes of the sample, and to expose and/or separate the strand(s) of the nucleic acid(s). This lysing and nucleic acid denaturing step to expose and separate the strands will allow amplification to occur much more readily.

The deoxyribonucleotide triphosphates dATP, dCTP, dGTP, and dTTP are added to the synthesis mixture, either separately or together with the primers, in adequate amounts and the resulting solution is heated to about 90°-100° C. from about 1 to 10 minutes, preferably from 1 to 4 minutes. After this heating period, the solution is allowed to cool, which is preferable for the primer hybridization. To the cooled mixture is added an appropriate agent for effecting the primer extension reaction (called herein "agent for polymerization"), and the reaction is allowed to occur under conditions known in the art. The agent for polymerization may also be added together with the other reagents if it is heat stable. This synthesis (or amplification) reaction may occur at room temperature up to a temperature above which the agent for polymerization no longer functions. Thus, for example, if DNA polymerase is used as the agent, the temperature is generally no greater than about 40° C. Most conveniently the reaction occurs at room temperature.

The allele specific oligonucleotide primers are useful in determining whether a subject is at risk of having breast or

5,654,155

5

ovarian cancer, and also useful for characterizing a tumor. Primers direct amplification of a target polynucleotide prior to sequencing. These unique BRCA1 oligonucleotide primers of TABLE II were designed and produced specifically to optimize amplification of portions of BRCA1 which are to be sequenced.

The primers used to carry out this invention embrace oligonucleotides of sufficient length and appropriate sequence to provide initiation of polymerization. Environmental conditions conducive to synthesis include the presence of nucleoside triphosphates and an agent for polymerization, such as DNA polymerase, and a suitable temperature and pH. The primer is preferably single stranded for maximum efficiency in amplification, but may be double stranded. If double stranded, the primer is first treated to separate its strands before being used to prepare extension products. The primer must be sufficiently long to prime the synthesis of extension products in the presence of the inducing agent for polymerization. The exact length of primer will depend on many factors, including temperature, buffer, and nucleotide composition. The oligonucleotide primer typically contains 12–20 or more nucleotides, although it may contain fewer nucleotides.

Primers used to carry out this invention are designed to be substantially complementary to each strand of the genomic locus to be amplified. This means that the primers must be sufficiently complementary to hybridize with their respective strands under conditions which allow the agent for polymerization to perform. In other words, the primers should have sufficient complementarity with the 5' and 3' sequences flanking the mutation to hybridize therewith and permit amplification of the genomic locus.

Oligonucleotide primers of the invention are employed in the amplification process which is an enzymatic chain reaction that produces exponential quantities of polymorphic locus relative to the number of reaction steps involved. Typically, one primer is complementary to the negative (–) strand of the polymorphic locus and the other is complementary to the positive (+) strand. Annealing the primers to denatured nucleic acid followed by extension with an enzyme, such as the large fragment of DNA polymerase I (Klenow) and nucleotides, results in newly synthesized + and – strands containing the target polymorphic locus sequence. Because these newly synthesized sequences are also templates, repeated cycles of denaturing, primer annealing, and extension results in exponential production of the region (i.e., the target polymorphic locus sequence) defined by the primers. The product of the chain reaction is a discrete nucleic acid duplex with termini corresponding to the ends of the specific primers employed.

The oligonucleotide primers of the invention may be prepared using any suitable method, such as conventional phosphotriester and phosphodiester methods or automated embodiments thereof. In one such automated embodiment, diethylphosphoramidites are used as starting materials and may be synthesized as described by Beaucage, et al., *Tetrahedron Letters*, 22:1859–1862, 1981. One method for synthesizing oligonucleotides on a modified solid support is described in U.S. Pat. No. 4,458,066.

The agent for polymerization may be any compound or system which will function to accomplish the synthesis of primer extension products, including enzymes. Suitable enzymes for this purpose include, for example, *E. coli* DNA polymerase I, Klenow fragment of *E. coli* DNA polymerase, polymerase mutants, reverse transcriptase, other enzymes, including heat-stable enzymes (e.i., those enzymes which

6

perform primer extension after being subjected to temperatures sufficiently elevated to cause denaturation), such as Taq polymerase. Suitable enzyme will facilitate combination of the nucleotides in the proper manner to form the primer extension products which are complementary to each polymorphic locus nucleic acid strand. Generally, the synthesis will be initiated at the 3' end of each primer and proceed in the 5' direction along the template strand, until synthesis terminates, producing molecules of different lengths.

The newly synthesized strand and its complementary nucleic acid strand will form a double-stranded molecule under hybridizing conditions described above and this hybrid is used in subsequent steps of the process. In the next step, the newly synthesized double-stranded molecule is subjected to denaturing conditions using any of the procedures described above to provide single-stranded molecules.

The steps of denaturing, annealing, and extension product synthesis can be repeated as often as needed to amplify the target polymorphic locus nucleic acid sequence to the extent necessary for detection. The amount of the specific nucleic acid sequence produced will accumulate in an exponential fashion.

Amplification is described in *PCR. A Practical Approach*, IRL Press, Eds. M. J. McPherson, P. Quirke, and G. R. Taylor, 1992.

The amplification products may be detected by Southern blots analysis, without using radioactive probes. In such a process, for example, a small sample of DNA containing a very low level of the nucleic acid sequence of the polymorphic locus is amplified, and analyzed via a Southern blotting technique or similarly, using dot blot analysis. The use of non-radioactive probes or labels is facilitated by the high level of the amplified signal. Alternatively, probes used to detect the amplified products can be directly or indirectly detectably labeled, for example, with a radioisotope, a fluorescent compound, a bioluminescent compound, a chemiluminescent compound, a metal chelator or an enzyme. Those of ordinary skill in the art will know of other suitable labels for binding to the probe, or will be able to ascertain such, using routine experimentation.

Sequences amplified by the methods of the invention can be further evaluated, detected, cloned, sequenced, and the like, either in solution or after binding to a solid support, by any method usually applied to the detection of a specific DNA sequence such as PCR, oligomer restriction (Saiki, et al., *Bio/Technology*, 3:1008–1012, 1985), allele-specific oligonucleotide (ASO) probe analysis (Conner, et al., *Proc. Natl. Acad. Sci. U.S.A.*, 80:278, 1983), oligonucleotide ligation assays (OLAs) (Landgren, et al., *Science*, 241:1007, 1988), and the like. Molecular techniques for DNA analysis have been reviewed (Landgren, et al., *Science*, 242:229–237, 1988).

Preferably, the method of amplifying is by PCR, as described herein and as is commonly used by those of ordinary skill in the art. Alternative methods of amplification have been described and can also be employed as long as the BRCA1 locus amplified by PCR using primers of the invention is similarly amplified by the alternative means. Such alternative amplification systems include but are not limited to self-sustained sequence replication, which begins with a short sequence of RNA of interest and a T7 promoter. Reverse transcriptase copies the RNA into cDNA and degrades the RNA, followed by reverse transcriptase polymerizing a second strand of DNA. Another nucleic acid amplification technique is nucleic acid sequence-based amplification (NASBA) which uses reverse transcription

A000110

and T7 RNA polymerase and incorporates two primers to target its cycling scheme. NASBA can begin with either DNA or RNA and finish with either, and amplifies to 10⁸ copies within 60 to 90 minutes. Alternatively, nucleic acid can be amplified by ligation activated transcription (LAT). LAT works from a single-stranded template with a single primer that is partially single-stranded and partially double-stranded. Amplification is initiated by ligating a cDNA to the promoter oligonucleotide and within a few hours, amplification is 10⁸ to 10⁹ fold. The QB replicase system can be utilized by attaching an RNA sequence called MDV-1 to RNA complementary to a DNA sequence of interest. Upon mixing with a sample, the hybrid RNA finds its complement among the specimen's mRNAs and binds, activating the replicase to copy the tag-along sequence of interest. Another nucleic acid amplification technique, ligase chain reaction (LCR), works by using two differently labeled halves of a sequence of interest which are covalently bonded by ligase in the presence of the contiguous sequence in a sample, forming a new target. The repair chain reaction (RCR) nucleic acid amplification technique uses two complementary and target-specific oligonucleotide probe pairs, thermostable polymerase and ligase, and DNA nucleotides to geometrically amplify targeted sequences. A 2-base gap separates the oligonucleotide probe pairs, and the RCR fills and joins the gap, mimicking normal DNA repair. Nucleic acid amplification by strand displacement activation (SDA) utilizes a short primer containing a recognition site for hincII with short overhang on the 5' end which binds to target DNA. A DNA polymerase fills in the part of the primer opposite the overhang with sulfur-containing adenine analogs. HincII is added but only cuts the unmodified DNA strand. A DNA polymerase that lacks 5' exonuclease activity enters at the cite of the nick and begins to polymerize, displacing the initial primer strand downstream and building a new one which serves as more primer. SDA produces greater than 10⁷-fold amplification in 2 hours at 37° C. Unlike PCR and LCR, SDA does not require instrumented Temperature cycling. Another amplification system useful in the method of the invention is the QB Replicase System. Although PCR is the preferred method of amplification if the invention, these other methods can also be used to amplify the BRCA1 locus as described in the method of the invention.

The BRCA1^(omi) Consensus Normal DNA sequence was obtained by end to end sequencing of five normal subjects in the manner described above followed by analysis of the data obtained. The data obtained provided us with the opportunity to evaluate six previously published normal polymorphisms for correctness and frequency in the population, and to identify an additional polymorphism not previously found.

The consensus normal gene sequence can be used for gene therapy. A complete description of the method is provided in Anderson et al. U.S. Pat. No. 5,399,346, issued Mar. 21, 1995. The isolated consensus normal BRCA1 gene can be constructed from amplification products and inserted into a vector such as the LXSXN vector. Fresh lymphocytes of a patient having a mutation in the BRCA1 gene, are cultured. The cells are transduced with the vector above, and culturing is continued. The cultured, transformed cells are infused into the patient.

The consensus normal BRCA1 amino acid sequence may be used to make diagnostic probes. Labeled diagnostic probes may be used by any hybridization method to determine the level of BRCA1 protein in serum or lysed cell suspension of a patient, or solid surface cell sample.

The consensus normal BRCA1 amino acid sequence may be used to provide a level of protection for patients against risk of breast or ovarian cancer. The isolated consensus normal BRCA1 gene can be constructed from amplification products and inserted into a vector such as the LXSXN vector. Fresh lymphocytes of a patient having a mutation in the BRCA1 gene, are cultured. The cells are transduced with the vector above, and cultured. Extracted BRCA1 protein can be provided by injection or other known means to patients who are at risk.

EXAMPLE 1

Determination of the Sequence of the BRCA1^(omi) Gene From Five Normal Individuals

Materials and Methods

Approximately 150 volunteers were screened in order to identify individuals with no cancer history in their immediate family (i.e. first and second degree relatives). Each person was asked to fill out a hereditary cancer prescreening questionnaire See TABLE I below. Five of these were randomly chosen for end-to-end sequencing of their BRCA1 gene." A first degree relative is a parent, sibling, or off spring. A second degree relative is an aunt, uncle, grandparent, grandchild, niece, nephew, or half-sibling.

TABLE I

Hereditary Cancer Pre-Screening Questionnaire

Part A: Answer the following questions about your family

1. To your knowledge, has anyone in your family been diagnosed with a very specific hereditary colon disease called Familial Adenomatous Polyposis (FAP)?

2. To your knowledge, have you or any aunt had breast cancer diagnosed before the age 35?

3. Have you had Inflammatory Bowel Disease, also called Crohn's Disease or Ulcerative Colitis, for more than 7 years?

Part B: Refer to the list of cancers below for your responses only to questions in Part B

Bladder Cancer	Lung Cancer	Pancreatic Cancer
Breast Cancer	Gastric Cancer	Prostate Cancer
Colon Cancer	Malignant Melanoma	Renal Cancer
Endometrial Cancer	Ovarian Cancer	Thyroid Cancer

4. Have your mother or father, your sisters or brothers or your children had any of the listed cancers?

5. Have there been diagnosed in your mother's brothers or sisters, or your mother's parents more than one of the cancers in the above list?

6. Have there been diagnosed in your father's brothers or sisters, or your father's parents more than one of the cancers in the above list?

Part C: Refer to the list of relatives below for responses only to questions in Part C

You	Your mother
Your sisters or brothers	Your mother's sisters or brothers (maternal aunts and uncles)
Your children	Your mother's parents (maternal grandparents)

7. Have there been diagnosed in these relatives 2 or more identical types of cancer?
Do not count "simple" skin cancer, also called basal cell or squamous cell skin cancer.

8. Is there a total of 4 or more of any cancers in the list of relatives above other than "simple" skin cancers?

Part D: Refer to the list of relatives below for responses only to questions in Part D.

You	Your father
Your sisters or brothers	Your father's sisters or brothers (paternal aunts and uncles)
Your children	Your father's parents (paternal grandparents)

5,654,155

9

10

TABLE I-continued

Hereditary Cancer Pre-Screening Questionnaire	
9.	Have there been diagnosed in these relatives 2 or more identical types of cancer? Do not count "simple" skin cancer, also called basal cell or squamous cell skin cancer.
10.	Is there a total of 4 or more of any cancers in the list of relatives above other than "simple" skin cancers?

© Copyright 1996, OncorMed, Inc.

Genomic DNA was isolated from white blood cells of five normal subjects selected from analysis of their answers to the questions above. Dideoxy sequence analysis was performed following polymerase chain reaction amplification.

All exons of the BRCA1 gene were subjected to direct dideoxy sequence analysis by asymmetric amplification using the polymerase chain reaction (PCR) to generate a single stranded product amplified from this DNA sample. Shuldiner, et al., Handbook of Techniques in Endocrine Research, p. 457-486, DePablo, F., Scanes, C., eds., Aca-

ademic Press, Inc., 1993. Fluorescent dye was attached for automated sequencing using the Taq Dye Terminator Kit (PERKIN-ELMER® cat# 401628). DNA sequencing was performed in both forward and reverse directions on an Applied Biosystems, Inc. (ABI) automated sequencer (Model 377). The software used for analysis of the resulting data was "Sequence Navigator" purchased through ABI.

1. Polymerase Chain Reaction (PCR) Amplification

Genomic DNA (100 nanograms) extracted from white blood cells of five normal subjects. Each of the five samples was sequenced end to end. Each sample was amplified in a final volume of 25 microliters containing 1 microliter (100 nanograms) genomic DNA, 2.5 microliters 10X PCR buffer (100 mM Tris, pH 8.3, 500 mM KCl, 1.2 mM MgCl₂), 2.5 microliters 10X dNTP mix (2 mM each nucleotide), 2.5 microliters forward primer, 2.5 microliters reverse primer, and 1 microliter Taq polymerase (5 units), and 13 microliters of water.

The primers in Table II, below were used to carry out amplification of the various sections of the BRCA1 gene samples. The primers were synthesized on an DNA/RNA Synthesizer Model 394®.

TABLE II

BRCA1 PRIMERS AND SEQUENCING DATA

EXON	SEQUENCE	SEQ. ID NO.	MER	MG ² + 1	SIZE
EXON 2	2F	5' GAA GTT GTC ATT TTA TAA ACC TTT-3'	3	24	1.6
	2R	5' TTT CTT TTC TTC CCT AGT ATG 5-3'	4	22	
EXON 3	3F	5' TCC TGA CAC AGC AGA CAT TTA-3'	5	21	1.4
	3R	5' TTG GAT TTT CGT TCT CAC TTA-3'	6	21	
EXON 5	5F	5' CTC TTA AGG GCA GTT GTG AG-3'	7	20	1.2
	5R	5' TTC CTA CTG TGG TTG CTT CC	8	20*	
EXON 6	6/7F	5' CTT ATT TTA GTG TCC TTA AAA GG-3'	9	23	1.6
	6R	5' TTT CAT GGA CAG CAC TTG AGT G-3'	10	22	
EXON 7	7F	5' CAC AAC AA GAG CAT ACA TAG GG-3'	11	23	1.5
	6/7R	5' TCG GGT TCA CTC TGT AGA AG-3'	12	20	
EXON 8	8F1	5' TTC TCT TCA GGA GGA AAA GCA-3'	13	21	1.2
	8R1	5' GCT GCC TAC CAC AAA TAC AAA-3'	14	21	
EXON 9	9F	5' CCA CAG TAG ATG CTC AGT AAATA-3'	15	23	1.2
	9R	5' TAG GAA AAT ACC AGC TTC ATA GA-3'	16	23	
EXON 10	10F	5' TGG TCA GCT TTC TGT AAT CG-3'	17	20	1.6
	10R	5' GTA TCT ACC CAC TCT CTT CAG-3'	18	24	
EXON 11A	11AF	5' CCA CCT CCA AGG TGT ATC A-3'	19	19	1.2
	11AR	5' TGT TAT GTT GGC TCC TTG CT-3'	20	20	
EXON 11B	11BF1	5' CAC TAA AGA CAG AAT GAA TCT A-3;	21	21	1.2
	11BR1	5' GAA GAA CCA GAA TAT TCA TCT A-3'	22	21	
EXON 11C	11CF1	5' TGA TGG GGA GTC TGA ATC AA-3'	23	20	1.2
	11CR1	5' TCT GCT TTC TTG ATA AAA TCC T-3'	24	22	
EXON 11D	11DF1	5' AGC GTC CCC TCA CAA ATA AA-3'	25	20	1.2
	11DR1	5' TCA AGC GCA TGA ATA TGC CT-3'	26	20	
EXON 11E	11EF	5' GTA TAA GCA ATA TGG AAC TCG A-3'	27	22	1.2
	11ER	5' TTA AGT TCA CTG GTA TTT GAA CA-3'	28	23	
EXON 11F	11FF	5' GAC AGC GAT ACT TTC CCA GA-3'	29	20	1.2
	11FR	5' TGG AAC AAC CAT GAA TTA GTC-3'	30	21	
EXON 11G	11GF	5' GGA AGT TAG CAC TCT AGG GA-3'	31	20	1.2
	11GR	5' GCA GTG ATA TTA ACT GTC TGT A-3'	32	22	
EXON 11H	11HF	5' TGG GTC CTT AAAGAA ACA AAGT-3'	33	22	1.2
	11HR	5' TCA GGT GAC ATT GAA TCT TCC-3'	34	21	
EXON 11I	11IF	5' CCA CTT TTT CCC ATC AAG TCA-3'	35	21	1.2
	11IR	5' TCA GGA TGC TTA CAA TTA CTT C-3'	36	21	
EXON 11J	11JF	5' CAA AAT TGA ATG CTA TGC TAA GA-3'	37	23	1.2
	11JR	5' TCG GTA ACC CTG AGC CAA AT-3'	38	20	
EXON 11K	11KF	5' GCA AAAGCG TCC AGA AAG GA-3'	39	20	1.2
	11KR-1	5' TAT TTG CAG TCA AGT CTT CCA A-3'	40	22	
EXON 11L	11LF-1	5' GTA ATA TTG GCA AAG GCA TCT-3'	41	22	1.2
	11LR	5' TAA AAT GTG CTC CCC AAA AGC A-3'	42	22	
EXON 12	12F	5' GTC CTG CCA ATG AGA AGA AA-3'	43	20	1.2
	12R	5' TGT CAG CAA ACC TAA GAA TGT-3'	44	21	
EXON 13	13F	5' AAT GGA AAG CTT CTC AAAGTA-3'	45	21	1.2
	13R	5' ATG TTG GAG CTA GGT CCT TAC-3'	46	21	
EXON 14	14F	5' CTA ACC TGA ATT ATC ACT ATC A-3'	47	22	1.2
	14R	5' GTG TAT AAATGC CTG TAT GCA-3'	48	21	

TABLE II-continued

BRCA1 PRIMERS AND SEQUENCING DATA						
EXON	SEQUENCE		SEQ. ID NO.	MER	MG ² + 1	SIZE
EXON 15	15F	5' TGG CTG CCC AGG AAG TAT G-3'	49	19	1.2	~375
	15R	5' AAC CAG AAT ATC TTT ATG TAG GA-3'	50	23		
EXON 16	16F	5' AAT TCT TAA CAG AGA CCA GAA C-3'	51	22	1.6	~550
	16R	5' AAA ACT CTT TCC AGA ATG TTG 5-3'	52	22		
EXON 17	17F	5' GTG TAG AAC GTG CAG GAT TG-3'	53	20	1.2	~275
	17R	5' TCG CCT CAT GTG GTT TTA-3'	54	18		
EXON 18	18F	5' GGC TCT TTA GCT TCT TAG GAC-3'	55	21	1.2	~350
	18R	5' GAG ACC ATT TTC CCA GCA TC-3'	56	20		
EXON 19	19F	5' CTG TCA TTC TTC CTG TGC TC-3'	57	20	1.2	~250
	19R	5' CAT TGT TAA GGA AAG TGG TGC-3'	58	21		
EXON 20	20F	5' ATA TGA CGT GTC TGC TCC AC-3'	59	20	1.2	~425
	20R	5' GGG AAT CCA AAT TAC ACA GC-3'	60	20		
EXON 21	21F	5' AAG CTC TTC CTT TTT GAA AGT C-3'	61	22	1.6	~300
	21R	5' GTA GAG AAA TAG AAT AGC CTC T-3'	62	22		
EXON 22	22F	5' TCC CAT TGA GAG GTC TTG CT-3'	63	20	1.6	~300
	22R	5' GAG AAG ACT TCT GAG GCT AC-3'	64	20		
EXON 23	23F-1	5' TGA AGT GAC AGT TCC AGT AGT-3'	65	21	1.2	~250
	23R-1	5' CAT TTT AGC CAT TCA TTC AAC AA-3'	66	23		
EXON 24	24F	5' ATG AAT TGA CAC TAA TCT CTG C-3'	67	22	1.4	~285
	24R	5' GTA GCC AGG ACA GTA GAA GGA-3'	68	2.1		

Thirty-five cycles were performed, each consisting of denaturing (95° C.; 30 seconds), annealing (55° C.; 1 minute), and extension (72° C.; 90 seconds), except during the first cycle in which the denaturing time was increased to 5 minutes, and during the last cycle in which the extension time was increased to 5 minutes.

PCR products were purified using QIA-QUICK® PCR purification kits (QIAGEN, cat# 28104; Chatsworth, Calif.). Yield and purity of the PCR product determined spectrophotometrically at OD₂₆₀ on a Beckman DU 650 spectrophotometer.

2. Dideoxy Sequence Analysis

Fluorescent dye was attached to PCR products for automated sequencing using the Taq Dye Terminator Kit (PERKIN-ELMER® cat#401628). DNA sequencing was performed in both forward and reverse directions on an Applied Biosystems, Inc. (ABI) Foster City, Calif., automated sequencer (Model 377). The software used for analysis of the resulting data was "SEQUENCE NAVIGATOR®" purchased through ABI.

3. Results

Differences in the nucleic acids of the five normal individuals were found in seven locations on the gene. The changes and their positions are found on TABLE III, below.

TABLE III

NORMAL PANEL TYPING							
AMINO ACID CHANGE	EXON	1	2	3	4	5	FREQUENCY
SER(SER) (694)	11E	C/C	C/T	C/T	T/T	T/T	0.4 C 0.6 T
LEU(LEU) (771)	11F	T/T	C/T	C/T	C/C	C/C	0.4 T 0.6 C
PRO(LEU) (871)	11G	C/T	C/T	C/T	T/T	T/T	0.3 C 0.7 T
GLU(GLY) (1038)	11I	A/A	A/G	A/G	G/G	G/G	0.4 A 0.6 G
LYS(ARG) (1183)	11J	A/A	A/G	A/G	G/G	G/G	0.4 A 0.6 G
SER(SER) (1436)	13	T/T	T/T	T/C	C/C	C/C	0.5 T 0.5 C
SER(GLY) (1613)	16	A/A	A/G	A/G	G/G	G/G	0.4 A 0.6 G

A consensus normal sequence of the BRCA1 gene was determined by calculating the percentage of occurrence of each polymorphism and inserting the more frequently occurring polymorphism into the published BRCA1 (Genbank Accession Number U15595). The normal consensus BRCA1^(omi) is shown as SEQ. ID NO. 1 at page 29.

The data show that for each of the samples. The BRCA1 gene is identical except in the region of seven polymorphisms. These polymorphic regions, together with their locations, the amino acid groups of each codon, the frequency of their occurrence and the amino acid coded for by each codon are found in TABLE IV below.

TABLE IV

CODON AND BASE CHANGES IN SEVEN POLYMORPHIC SITES OF BRCA1 NORMAL GENE							
SAMPLE NUMBER	BASE CHANGE	POSITION	EXON	CODON CHANGE	AA CHANGE	PUBLISHED FREQUENCY	² REFERENCES
2, 3, 4, 5	C-T	2201	11E	AGC(AGT)	SER—SER (694)	C	UNPUBLISHED
2, 3, 4, 5	T-C	2430	11F	TTG(CTG)	LEU—LEU (771)	T = 67%	13
1, 2, 3, 4, 5	C-T	2731	11G	CCG(CTG)	PRO—LEU (871)	T = 66%	12
2, 3, 4, 5	A-G	3232	11I	GAA(GGA)	GLU—GLY (1038)	A = 67%	13
2, 3, 4, 5	A-G	3667	11J	AAA(AGA)	LYS—ARG (1183)	A = 68%	12
3, 4, 5	T-C	4427	13	TCT(TCC)	SER—SER (1436)	T = 67%	12
2, 3, 4, 5	A-G	4956	16	AGT(GGT)	SER—GLY (1613)	A = 67%	12

²Reference numbers correspond To The Table Of References on Page 28.

EXAMPLE 2

Determination of a Normal Individual Using BRCA1^(omi) and the Seven Polymorphisms for Reference

A person skilled in the art of genetic susceptibility testing will find the present invention useful for:

- a) identifying individuals having a normal BRCA1 gene, who are therefore not at risk or have no increased susceptibility to breast or ovarian cancer from a BRCA1 mutation;
- b) avoiding misinterpretation of normal polymorphisms found in the normal BRCA1 gene;

Sequencing is carried out as in EXAMPLE 1 using a blood sample from the patient in question. However, the BRCA1^(omi) sequence is used for reference and polymorphic sites are compared to the nucleic acid sequences listed above for normal codons at each polymorphic site. A normal sample is one which compares to the BRCA1^(omi) sequence and contains one of the normal base variations which occur at each of the polymorphic sites. The codons which occur at each of the polymorphic sites are paired below.

- AGC and AGT at position 2201,
- TTG and CTG at position 2430,
- CCG and CTG at position 2731,
- GAA and GGA at position 3232,
- AAA and AGA at position 3667,
- TCT and TCC at position 4427, and
- AGT and GGT at position 4956.

The availability of these polymorphic pairs provides added assurance that one skilled in the art can correctly interpret the polymorphic variations without mistaking a normal variation for a mutation.

Exon 11 of the BRCA1 gene is subjected to direct dideoxy sequence analysis by asymmetric amplification using the polymerase chain reaction (PCR) to generate a single stranded product amplified from this DNA sample. Shuldiner, et al., Handbook of Techniques in Endocrine Research, p. 457–486, DePablo, F., Scanes, C., eds., Academic Press, Inc., 1993. Fluorescent dye is attached for automated sequencing using the Taq Dye Terminator Kit (PERKIN-ELMER® cat# 401628). DNA sequencing is performed in both forward and reverse directions on an Applied Biosystems, Inc. (ABI) automated Sequencer (Model 377).

The software used for analysis of the resulting data is “SEQUENCE NAVIGATOR®” purchased through ABL.

1. Polymer Chain Reaction (PCR) Amplification

- 25 Genomic DNA (100 nanograms) extracted from white blood cells of the subject is amplified in a final volume of 25 microliters containing 1 microliter (100 nanograms) genomic DNA, 2.5 microliters 10X PCR buffer (100 mM Tris, pH 8.3, 500 mM KCl, 1.2 mM MgCl₂), 2.5 microliters 10X dNTP mix (2 mM each nucleotide), 2.5 microliters forward primer (BRCA1-11K-F, 10 micromolar solution), 2.5 microliters reverse primer (BRCA1-11K-R, 10 micromolar solution), and 1 microliter Taq polymerase (5 units), and 13 microliters of water. The PCR primers used to amplify a patient’s sample BRCA1 gene are listed in Table II. The primers were synthesized on an DNA/RNA Synthesizer Model 394®. Thirty-five cycles are of amplification are performed, each consisting of denaturing (95° C.; 30 seconds), annealing (55° C.; 1 minute), and extension (72° C.; 90 seconds), except during the first cycle in which the denaturing time is increased to 5 minutes, and during the last cycle in which the extension time is increased to 5 minutes.

- 35 PCR products are purified using QIA-QUICK® PCR purification kits (QIAGEN®, cat# 28104; Chatsworth, Calif.). Yield and purity of the PCR product determined spectrophotometrically at OD₂₆₀ on a Beckman DU 650 spectrophotometer.

2. Dideoxy Sequence Analysis

- 50 Fluorescent dye is attached to PCR products for automated sequencing using the Taq Dye Terminator Kit (PERKIN-ELMER® cat# 401628). DNA sequencing is performed in both forward and reverse directions on an Applied Biosystems, Inc. (ABI) Foster City, Calif., automated sequencer (Model 377). The software used for analysis of the resulting data is “SEQUENCE NAVIGATOR®” purchased through ABL. The BRCA1^(omi) sequence is entered into the Sequence Navigator software as the Standard for comparison. The Sequence Navigator software compares the sample sequence to the BRCA1^(omi) standard, base by base. 60 The Sequence Navigator highlights all differences between the BRCA1^(omi) (SEQ ID. NO: 1) consensus normal DNA sequence and the patient’s sample sequence.

- 65 A first technologist checks the computerized results by comparing visually the BRCA1^(omi) standard against the patient’s sample, and again highlights any differences between the standard and the sample. The first primary technologist then interprets the sequence variations at each

5,654,155

15

position along the sequence. Chromatograms from each sequence variation are generated by the Sequence Navigator and printed on a color printer. The peaks are interpreted by the first primary technologist and a second primary technologist. A secondary technologist then reviews the chromatograms. The results are finally interpreted by a geneticist. In each instance, a variation is compared to known normal polymorphisms for position and base change. If the sample BRCA1 sequence matches the BRCA1^(omi) standard, with only variations within the known list of polymorphisms, it is interpreted as a normal gene sequence.

EXAMPLE 3

Determining the Presence of a Mutation in the BRCA1 Gene Using BRCA1^(omi) and Seven Polymorphisms for Reference

A person skilled in the art of genetic susceptibility testing will find the present invention useful for determining the presence of a known or previously unknown mutation in the BRCA1 gene. A list of mutations of BRCA1 is publicly available in the Breast Cancer Information Core at: http://www.ncbi.nlm.nih.gov/ncic/ncic_lab_transfer/bic. This data site became publicly available on Nov. 1, 1995. Friend, S. et al. *Nature Genetics* 11:238, (1995). Sequencing is carried out as in EXAMPLE 1 using a blood sample from the patient in question. However, the BRCA1^(omi) (SEQ ID NO: 1) sequence is used for reference and polymorphic sites are compared to the nucleic acid sequences listed above for normal codons at each polymorphic site. A normal sample is one which compares to the BRCA1^(omi) sequence and contains one of the normal base variations which occur at each of the polymorphic sites. The codons which occur at each of the polymorphic sites are paired below.

- AGC and AGT at position 2201,
- TTG and CTG at position 2430,
- CCG and CTG at position 2731,
- GAA and GGA at position 3232,
- AAA and AGA at position 3667,
- TCT and TCC at position 4427, and
- AGT and GGT at position 4956.

The availability of these polymorphic pairs provides added assurance that one skilled in the art can correctly interpret the polymorphic variations without mistaking a normal variation for a mutation.

As evident from the data in Table III and IV, the statistical analysis (on average) shows one or more normal codon pairs wherein the codons occur in the following frequencies in the normal population, respectively:

- at position 2201, AGC and AGT occur at frequencies of about 40%, and from about 55–65%, respectively;
- at position 2430, TTG and CTG occur at frequencies from about 35–45%, and from about 55–65%, respectively;
- at position 2731, CCG and CTG occur at frequencies from about 25–35%, and from about 65–75%, respectively;
- at position 3232, GAA and GGA occur at frequencies from about 35–45%, and from about 55–65%, respectively;
- at position 3667, AAA and AGA occur at frequencies from about 35–45%, and from about 55–65%, respectively;
- at position 4427, TCT and TCC occur at frequencies from about 45–55%, and from about 45–55%, respectively;
- and

16

at position 4956, AGT and GGT occur at frequencies from about 35–45%, and from about 55–65%, respectively.

Exon 11 of the BRCA1 gene is subjected to direct dideoxy sequence analysis by asymmetric amplification using the polymerase chain reaction (PCR) to generate a single stranded product amplified from this DNA sample. Shuldiner, et al., *Handbook of Techniques in Endocrine Research*, p. 457–486, DePablo, F., Scanes, C., eds., Academic Press, Inc., 1993. Fluorescent dye is attached for automated sequencing using the Taq Dye Terminator Kit (PERKIN-ELMER® cat# 401628). DNA sequencing is performed in both forward and reverse directions on an Applied Biosystems, Inc. (ABI) automated sequencer (Model 377). The software used for analysis of the resulting data is “SEQUENCE NAVIGATOR®” purchased through ABI.

1. Polymerase Chain Reaction (PCR) Amplification

Genomic DNA (100 nanograms) extracted from white blood cells of the subject is amplified in a final volume of 25 microliters containing 1 microliter (100 nanograms) genomic DNA, 2.5 microliters 10X PCR buffer (100 mM Tris, pH 8.3, mM KCl, 1.2 mM MgCl₂), 2.5 microliters 10X dNTP mix (2 mM each nucleotide), 2.5 microliters forward primer (BRCA1-11K-F, 10 micromolar solution), 2.5 microliters reverse primer (BRCA1-11K-R, 10 micromolar solution), and 1 microliter Taq polymerase (5 units), and 13 microliters of water. The PCR primers used to amplify a patient's sample BRCA1 gene are listed in Table II. The primers were synthesized on an DNA/RNA Synthesizer Model 394®. Thirty-five cycles are of amplification are performed, each consisting of denaturing (95° C. 30 seconds), annealing (55° C. 1 minute), and extension (72° C.; 90 seconds), except during the first cycle in which the denaturing time is increased to 5 minutes, and during the last cycle in which the extension time is increased to 5 minutes.

PCR products are purified using QIA-QUICK® PCR purification kits (QIAGEN®, cat# 28104; Chatsworth, Calif.). Yield and purity of the PCR product determined spectrophotometrically at OD₂₆₀ on a Beckman DU 650 spectrophotometer.

2. Dideoxy Sequence Analysis

Fluorescent dye is attached to PCR products for automated sequencing using the Taq Dye Terminator Kit (PERKIN-ELMER® cat# 401628). DNA sequencing is performed in both forward and reverse directions on an Applied Biosystems, Inc. (ABI) Foster City, Calif., automated sequencer (Model 377). The software used for analysis of the resulting data is “SEQUENCE NAVIGATOR®” purchased through ABI. The BRCA1^(omi) sequence (SEQ ID. NO: 1) is entered into the Sequence Navigator software as the Standard for comparison. The Sequence Navigator software compares the sample sequence to the BRCA1^(omi) standard (SEQ ID NO: 1), base by base. The Sequence Navigator highlights all differences between the BRCA1^(omi) consensus normal DNA sequence and the patient's sample sequence.

A first technologist checks the computerized results by comparing visually the BRCA1^(omi) standard (SEQ ID. NO: 1) against the patient's sample, and again highlights any differences between the standard and the sample. The first primary technologist then interprets the sequence variations at each position along the sequence. Chromatograms from each sequence variation are generated by the Sequence Navigator and printed on a color printer. The peaks are interpreted by the first primary technologist and also by a second primary technologist. A secondary technologist then reviews the chromatograms. The results are finally interpreted by a geneticist. In each instance, a variation is

compared to known normal polymorphisms for position and base change. If the sample BRCA1 sequence matches the BRCA1^(omi) standard, with only variations within the known list of polymorphisms, it is interpreted as a normal gene sequence. Mutations are noted by the length of non-matching variation. Such a lengthy mismatch pattern occurs with deletions and substitutions. An alteration such as a base substitution at a single position will be noted as a single mismatch between the standard and the patient's gene sample.

EXAMPLE 4

Determining the Presence of a Mutation in the BRCA1 Gene Using BRCA1^(omi) and Seven Polymorphisms for Reference

A person skilled in the art of genetic susceptibility testing will find the present invention useful for determining the presence of a known or previously unknown mutation in the BRCA1 gene. A list of mutations of BRCA1 is publicly available in the Breast Cancer Information Core at: http://www.ncbi.nlm.nih.gov/DIR/lab_transfer/bic. This data site became publicly available on Nov. 1, 1995. Friend, S. et al. *Nature Genetics* 11:238, (1995). In this example, a mutation in exon 11 is characterized by amplifying the region of the mutation with a primer which matches the region of the mutation. Exon 11 of the BRCA1 gene is subjected to direct dideoxy sequence analysis by asymmetric amplification using the polymerase chain reaction (PCR) to generate a single stranded product amplified from this DNA sample. Shuldiner, et al., *Handbook of Techniques in Endocrine Research*, p. 457-486, DePablo, F., Scanes, C., eds., Academic Press, Inc., 1993. Fluorescent dye is attached for automated sequencing using the Taq Dye Terminator Kit (Perkin-Elmer® cat# 401628). DNA sequencing is performed in both forward and reverse directions on an Applied Biosystems, Inc. (ABI) automated sequencer (Model 377). The software used for analysis of the resulting data is "Sequence Navigator" purchased through ABI. 1. Polymerase Chain Reaction (PCR) Amplification Genomic DNA (100 nanograms) extracted from white blood cells of the subject is amplified in a final volume of 25 microliters containing 1 microliter (100 nanograms) genomic DNA, 2.5 microliters 10X PCR buffer (100 mM Tris, pH 8.3, 500 mM KCl, 1.2 mM MgCl₂), 2.5 microliters 10X dNTP mix (2 mM each nucleotide), 2.5 microliters forward primer (BRCA1-11K-F, 10 micromolar solution), 2.5 microliters reverse primer (BRCA1-11K-R, 10 micromolar solution), and 1 microliter Taq polymerase (5 units), and 13 microliters of water. The PCR primers used to amplify segment K of exon 11 (where the mutation is found) are as follows:

BRCA1-11K-F: 5'-GCA AAA GCG TCC AGA AAG GA-3' SEQ ID NO: 69
BRCA1-11K-R: 5'-AGT CTT CCAATT CACTGCAC-3' SEQ ID NO: 70

The primers are synthesized on an DNA/RNA Synthesizer Model 394®. Thirty-five cycles are performed, each consisting of denaturing (95° C. 30 seconds), annealing (55° C.; 1 minute), and extension (72° C.; 90 seconds), except during the first cycle in which the denaturing time is increased to 5 minutes, and during the last cycle in which the extension time is increased to 5 minutes. PCR products are purified using QIA-QUICK® PCR purification kits (Qiagen®, cat# 28104; Chatsworth, Calif.). Yield and purity of the PCR product determined spectrophotometrically at OD₂₆₀ on a Beckman DU 650 spectrophotometer.

2. Dideoxy Sequence Analysis

Fluorescent dye is attached to PCR products for automated sequencing using the Taq Dye Terminator Kit (PERKIN-ELMER® Cat # 401628). DNA sequencing is performed in both forward and reverse directions on an Applied Biosystems, Inc. (ABI) Foster City, Calif., automated sequencer (Model 377). The software used for analysis of the resulting data is "SEQUENCE NAVIGATOR" purchased through ABI. The BRCA1^(omi) sequence is entered into the SEQUENCE Navigator software as the Standard for comparison. The Sequence Navigator software compares the sample sequence to the BRCA1^(omi) standard (SEQ ID. NO: 1), base by base. The Sequence Navigator highlights all differences between the BRCA1^(omi) (SEQ ID. NO: 1) consensus normal DNA sequence and the patient's sample sequence.

A first technologist checks the computerized results by comparing visually the BRCA1^(omi) standard against the patient's sample, and again highlights any differences between the standard and the sample. The first primary technologist then interprets the sequence variations at each position along the sequence. Chromatograms from each sequence variation are generated by the Sequence Navigator and printed on a color printer. The peaks are interpreted by the first primary technologist and a second primary technologist. A secondary technologist then reviews the chromatograms. The results are finally interpreted by a geneticist. In each instance, a variation is compared to known normal polymorphisms for position and base change. The seven known polymorphisms which occur in the Consensus Normal DNA sequence are:

AGC and AGT at position 2201,
TTG and CTG at position 2430,
CCG and CTG at position 2731,
GAA and GGA at position 3232,
AAA and AGA at position 3667,
TCT and TCC at position 4427, and
AGT and GGT at position 4956.

Mutations are noted by the length of non-matching variation. Such a lengthy mismatch pattern occurs with deletions and substitutions.

5. Result

Using the above PCR amplification and standard fluorescent sequencing technology, The 3888delGA mutation may be found. The 3888delGA mutation The BRCA1 gene lies in segment "K" of exon 11. The DNA sequence results demonstrate the presence of a two base pair deletion at nucleotides 3888 and 3889 of the published BRCA1^(omi) sequence. This mutation interrupts the normal reading frame of the BRCA1 transcript, resulting in the appearance of an in-frame terminator (TAG) at codon position 1265. This mutation is, therefore, predicted to result in a truncated, and most likely, non-functional protein. The formal name of the mutation will be 3888delGA. This mutation is named in accordance with the suggested nomenclature for naming mutations, Baudet, A et al., *Human Mutation* 2:245-248, (1993).

TABLE OF REFERENCES

1. Sanger, F., et al., *J. Mol. Biol.* 42:1617, (1980).
2. Beaucage, et al., *Tetrahedron Letters* 22:1859-1862, (1981).
3. Maniatis, et al. in *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor, NY, p 280-281, (1982).
4. Conner, et al., *Proc. Natl. Acad. Sci. U.S.A.* 80:278, (1983)
5. Saiki, et al., *BioTechnology* 3:1008-1012, (1985)

5,654,155

19

-continued

TABLE OF REFERENCES

6. Landgren, et. al., Science 241:1007, (1988)
7. Landgren, et. al., Science 242:229-237, (1988).
8. PCR. A Practical Approach, IRL Press, Eds. M. J. McPherson, P. Quirke, and G. R. Taylor, (1992).
9. Easton et al., American Journal of Human Genetics 52:678-701, (1993).
10. U.S. Pat. No. 4,458,066.
11. Rowell, S., et al., American Journal of Human Genetics 55:861-865, (1994)
12. Miki, Y. et al., Science 266:66-71, (1994).
13. Friedman, L. et al., Nature Genetics 8:399-404, (1994).

20

-continued

TABLE OF REFERENCES

5 14. Baudet, A et al., Human Mutation 2:245-248, (1993).
15. Friend, S. et al. Nature Genetics 11:238, (1995).

10 Although the invention has been described with reference to the presently preferred embodiments, it should be understood that various modifications can be made without departing from the spirit of the invention. Accordingly, the invention is limited only by the following claims.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i i i) NUMBER OF SEQUENCES: 74

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5711 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: Not Relevant
- (D) TOPOLOGY: linear

(i i) MOLECULE TYPE: cDNA

(v i) ORIGINAL SOURCE:

- (A) ORGANISM: Homo sapiens
- (B) STRAIN: BRCA1

(v i i i) POSITION IN GENOME:

- (A) CHROMOSOME/SEGMENT: 17
- (B) MAP POSITION: 17q21

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:1:

AGCTCGCTGA GACTTCCTGG ACCCCGCACC AGGCTGTGGG GTTCTCAGA TAACTGGGCC 60
CCTGCGCTCA GGAGGCCCTC ACCCTCTGCT CTGGGTAAAG TTCATTGGAA CAGAAAGAAA 120
TGGATTTATC TGCTCTTCGC GTTGAAGAAG TACAAAATGT CATTAATGCT ATGCAGAAAA 180
TCTTAGAGTG TCCCATCTGT CTGGAGTTGA TCAAGGAACC TGTCTCCACA AAGTGTGACC 240
ACATATTTTG CAAATTTTGC ATGCTGAAAC TTCTCAACCA GAAGAAAGGG CCTTCACAGT 300
GTCCTTTATG TAAGAATGAT ATAACCAAAA GGAGCCTACA AGAAAGTACG AGATTTAGTC 360
AACTTGTTGA AGAGCTATTG AAAATCATTT GTGCTTTTCA GCTTGACACA GGTTTGGAGT 420
ATGCAAACAG CTATAATTTT GCAAAAAAGG AAAATAACTC TCCTGAACAT CTAAAAGATG 480
AAGTTTCTAT CATCCAAAGT ATGGGCTACA GAAACCGTGC CAAAAGACTT CTACAGAGTG 540
AACCCGAAAA TCCTTCCTTG CAGGAAACCA GTCTCAGTGT CCAACTCTCT AACCTTGAA 600
CTGTGAGAAC TCTGAGGACA AAGCAGCGGA TACAACCTCA AAAGACGTCT GTCTACATTG 660
AATTGGGATC TGATTCTTCT GAAGATACCG TTAATAAGGC AACTTATTGC AGTGTGGGAG 720
ATCAAGAATT GTTACAAATC ACCCCTCAAG GAACCAGGGA TGAAATCAGT TTGGATTCTG 780
CAAAAAAGGC TGCTTGTGAA TTTTCTGAGA CGGATGTAAAC AAATACTGAA CATCATCAAC 840
CCAGTAATAA TGATTTGAAC ACCACTGAGA AGCGTGCAGC TGAGAGGCAT CCAGAAAAGT 900
ATCAGGGTAG TTCTGTTTCA AACTTGCAATG TGGAGCCATG TGGCACAAAT ACTCATGCCA 960
GCTCATTACA GCATGAGAAC AGCAGTTTAT TACTACTAA AGACAGAATG AATGTAGAAA 1020

21			5,654,155			22		
-continued								
AGGCTGAATT	CTGTAATAAA	AGCAAACAGC	CTGGCTTAGC	AAGGAGCCAA	CATAACAGAT	1080		
GGGCTGGAAG	TAAGGAAACA	TGTAATGATA	GGCGGACTCC	CAGCACAGAA	AAAAAGGTAG	1140		
ATCTGAATGC	TGATCCCCTG	TGTGAGAGAA	AAGAATGGAA	TAAGCAGAAA	CTGCCATGCT	1200		
CAGAGAATCC	TAGAGATACT	GAAGATGTTC	CTTGGATAAC	ACTAAATAGC	AGCATTTCAGA	1260		
AAGTTAATGA	GTGGTTTTTC	AGAAGTGATG	AACGTGTTAGG	TTCTGATGAC	TCACATGATG	1320		
GGGAGTCTGA	ATCAAATGCC	AAAGTAGCTG	ATGTATTGGA	CGTTCTAAAT	GAGGTAGATG	1380		
AATATTCTGG	TTCTTCAGAG	AAAATAGACT	TACTGGCCAG	TGATCCTCAT	GAGGCTTTAA	1440		
TATGTAAAAAG	TGAAAGAGTT	CACTCCAAAT	CAGTAGAGAG	TAATATTGAA	GACAAAATAT	1500		
TTGGGAAAAC	CTATCGGAAG	AAGGCAAGCC	TCCCCAACTT	AAGCCATGTA	ACTGAAAATC	1560		
TAATTATAGG	AGCATTGTGT	ACTGAGCCAC	AGATAATACA	AGAGCGTCCC	CTCACAAATA	1620		
AATTAAGGCG	TAAAAGGAGA	CCTACATCAG	GCCTTCATCC	TGAGGATTTT	ATCAAGAAAG	1680		
CAGATTTGGC	AGTTCAAAAAG	ACTCCTGAAA	TGATAAATCA	GGGAACATAAC	CAAACGGAGC	1740		
AGAATGGTCA	AGTGATGAAT	ATTACTAATA	GTGGTCATGA	GAATAAAAACA	AAAGGTGATT	1800		
CTATTTCAGAA	TGAGAAAAAT	CCTAACCCAA	TAGAATCACT	CGAAAAAGAA	TCTGCTTTCA	1860		
AAACGAAAAGC	TGAACCTATA	AGCAGCAGTA	TAAGCAATAT	GGAACCTCGAA	TTAAATATCC	1920		
ACAATTCAAA	AGCACCTAAA	AAGAATAGGC	TGAGGAGGAA	GTCTTCTACC	AGGCATATTC	1980		
ATGCGCTTGA	ACTAGTAGTC	AGTAGAAAATC	TAAGCCCACC	TAATTGTACT	GAATTGCAAA	2040		
TTGATAGTTG	TTCTAGCAGT	GAAGAGATAA	AGAAAAAATA	GTACAACCAA	ATGCCAGTCA	2100		
GGCACAGCAG	AAACCTACAA	CTCATGGAAG	GTAAAGAACC	TGCAACTGGA	GCCAAGAAGA	2160		
GTAACAAGCC	AAATGAACAG	ACAAGTAAAA	GACATGACAG	TGATACTTTC	CCAGAGCTGA	2220		
AGTTAACAAA	TGCACCTGGT	TCTTTTACTA	AGTGTTCAAA	TACCAGTGAA	CTTAAAGAAT	2280		
TTGTCAATCC	TAGCCTTCCA	AGAGAAGAAA	AAGAAGAGAA	ACTAGAAACA	GTTAAAGTGT	2340		
CTAATAATGC	TGAAGACCCC	AAAGATCTCA	TGTTAAGTGG	AGAAAGGGTT	TTGCAAACTG	2400		
AAAGATCTGT	AGAGAGTAGC	AGTATTTTAC	TGGTACCTGG	TACTGATTAT	GGCACTCAGG	2460		
AAAGTATCTC	GTTACTGGAA	GTTAGCACTC	TAGGGAAGGC	AAAAACAGAA	CCAAATAAAT	2520		
GTGTGAGTCA	GTGTGCAGCA	TTTGAAAACC	CCAAGGGACT	AATTCATGGT	TGTTCCAAAG	2580		
ATAATAGAAA	TGACACAGAA	GGCTTTAAGT	ATCCATTGGG	ACATGAAAGT	AACCACAGTC	2640		
GGGAAACAAG	CATAGAAATG	GAAGAAAAGTG	AAC TTGATGC	TCAGTATTTG	CAGAATACAT	2700		
TCAAGGTTTC	AAAGCGCCAG	TCATTTGCTC	TGTTTTCAAA	TCCAGGAAAT	GCAGAAGAGG	2760		
AATGTGCAAC	ATTCTCTGCC	CACTCTGGGT	CCTTAAAGAA	ACAAAGTCCA	AAAGTCACTT	2820		
TTGAATGTGA	ACAAAAGGAA	GAAAATCAAG	GAAAGAATGA	GTCTAATATC	AAGCCTGTAC	2880		
AGACAGTTAA	TATCACTGCA	GGCTTTCCTG	TGGTTGGTCA	GAAAGATAAG	CCAGTTGATA	2940		
ATGCCAAATG	TAGTATCAAA	GGAGGCTCTA	GGTTTTGTCT	ATCATCTCAG	TTCAGAGGCA	3000		
ACGAAACTGG	ACTCATTACT	CCAAATAAAC	ATGGACTTTT	ACAAAACCCA	TATCGTATAC	3060		
CACCACTTTT	TCCCATCAAG	TCATTTGTTA	AAACTAAATG	TAAGAAAAAT	CTGCTAGAGG	3120		
AAAACTTTGA	GGAACATTCA	ATGTCACCTG	AAAGAGAAAT	GGGAAATGAG	AACATTCCAA	3180		
GTACAGTGAG	CACAATTAGC	CGTAATAACA	TTAGAGAAAA	TGTTTTTAAA	GGAGCCAGCT	3240		
CAAGCAATAT	TAATGAAGTA	GGTTCCAGTA	CTAATGAAGT	GGGCTCCAGT	ATTAATGAAA	3300		
TAGGTTCCAG	TGATGAAAAC	ATTCAAGCAG	AAC TAGGTAG	AAACAGAGGG	CCAAAAATTGA	3360		
ATGCTATGCT	TAGATTAGGG	GTTTTGCAAC	CTGAGGTCTA	TAAACAAAGT	CTTCTTGAA	3420		

23		5,654,155		24		
-continued						
GTAATTGTAA	GCATCCTGAA	ATAAAAAAGC	AAGAATATGA	AGAAGTAGTT	CAGACTGTTA	3480
ATACAGATTT	CTCTCCATAT	CTGATTTCAG	ATAACTTAGA	ACAGCCTATG	GGAAGTAGTC	3540
ATGCATCTCA	GGTTTGTTC	GAGACACCTG	ATGACCTGTT	AGATGATGGT	GAAATAAAGG	3600
AAGATACTAG	TTTTGCTGAA	AATGACATTA	AGGAAAGTTC	TGCTGTTTTT	AGCAAAAGCG	3660
TCCAGAGAGG	AGAGCTTAGC	AGGAGTCCTA	GCCCTTTCAC	CCATACACAT	TTGGCTCAGG	3720
GTTACCGAAG	AGGGGCCAAG	AAATTAGAGT	CCTCAGAAGA	GAACCTATCT	AGTGAGGATG	3780
AAGAGCTTCC	CTGCTTCCAA	CACTTGTTAT	TTGGTAAAGT	AAACAATATA	CCTTCTCAGT	3840
CTACTAGGCA	TAGCACC GTT	GCTACCGAGT	GTCIGTCTAA	GAACACAGAG	GAGAATTTAT	3900
TATCATTGAA	GAATAGCTTA	AATGACTGCA	GTAACCAGGT	AATATTGGCA	AAGGCATCTC	3960
AGGAACATCA	CCTTAGTGAG	GAAACAAAAT	GTTCTGCTAG	CTTGTTTTCT	TCACAGTGCA	4020
GTGAATTGGA	AGACTTGACT	GCAAATACAA	ACACCCAGGA	TCCITTTCTTG	ATTGGTTCTT	4080
CCAAACAAAT	GAGGCATCAG	TCTGAAAGCC	AGGGAGTTGG	TCTGAGTGAC	AAGGAATTGG	4140
TTTCAGATGA	TGAAGAAAGA	GGAACGGGCT	TGGAAGAAAA	TAATCAAGAA	GAGCAAAGCA	4200
TGGATTCAAA	CTTAGGTGAA	GCAGCATCTG	GGTGTGAGAG	TGAAACAAGC	GTCTCTGAAG	4260
ACTGCTCAGG	GCTATCCTCT	CAGAGTGACA	TTTTAACCAC	TCAGCAGAGG	GATACCATGC	4320
AACATAACCT	GATAAAGCTC	CAGCAGGAAA	TGGCTGAACT	AGAAGCTGTG	TTAGAACAGC	4380
ATGGGAGCCA	GCCTTCTAAC	AGCTACCCTT	CCATCATAAG	TGACTCCTCT	GCCCTTGAGG	4440
ACCTGCGAAA	TCCAGAACAA	AGCACATCAG	AAAAAGCAGT	ATTAACCTCA	CAGAAAAGTA	4500
GTGAATACCC	TATAAGCCAG	AATCCAGAAG	GCCTTTCTGC	TGACAAGTTT	GAGGTGTCTG	4560
CAGATAGTTC	TACCAGTAAA	AATAAAGAAC	CAGGAGTGGA	AAGGTCATCC	CCTTCTAAAT	4620
GCCCATCATT	AGATGATAGG	TGGTACATGC	ACAGTTGCTC	TGGGAGTCTT	CAGAATAGAA	4680
ACTACCCATC	TCAAGAGGAG	CTCATTAAGG	TTGTTGATGT	GGAGGAGCAA	CAGCTGGAAG	4740
AGTCTGGGCC	ACACGATTTG	ACGGAAACAT	CTTACTTGCC	AAGGCAAGAT	CTAGAGGGAA	4800
CCCCTTACCT	GGAATCTGGA	ATCAGCCTCT	TCTCTGATGA	CCCTGAATCT	GATCCTTCTG	4860
AAGACAGAGC	CCCAGAGTCA	GCTCGTGTTG	GCAACATACC	ATCTTCAACC	TCTGCATTGA	4920
AAGTTCCCCA	ATTGAAAAGT	GCAGAATCTG	CCCAGGGTCC	AGCTGCTGCT	CATACTACTG	4980
ATACTGCTGG	GTATAATGCA	ATGGAAGAAA	GTGTGAGCAG	GGAGAAGCCA	GAATTGACAG	5040
CTTCAACAGA	AAGGGTCAAC	AAAAGAATGT	CCATGGTGGT	GTCTGGCCTG	ACCCCAGAAG	5100
AATTTATGCT	CGTGTACAAG	TTTGCCAGAA	AACACCACAT	CACTTTAACT	AATCTAATTA	5160
CTGAAGAGAC	TACTCATGTT	GTTATGAAAA	CAGATGCTGA	GTTTGTGTGT	GAACGGACAC	5220
TGAAATATTT	TCTAGGAATT	GCGGGAGGAA	AATGGGTAGT	TAGCTATTTT	TGGGTGACCC	5280
AGTCTATTAA	AGAAAGAAAA	ATGCTGAATG	AGCATGATTT	TGAAGTCAGA	GGAGATGTGG	5340
TCAATGGAAG	AAACCACCAA	GGTCCAAAGC	GAGCAAGAGA	ATCCCAGGAC	AGAAAGATCT	5400
TCAGGGGGCT	AGAAATCTGT	TGCTATGGGC	CCTTCACCAA	CATGCCCACA	GATCAACTGG	5460
AATGGATGGT	ACAGCTGTGT	GGTGCTTCTG	TGGTGAAGGA	GCTTTCATCA	TTCACCCTTG	5520
GCACAGGTGT	CCACCCAATT	GTGGTTGTGC	AGCCAGATGC	CTGGACAGAG	GACAATGGCT	5580
TCCATGCAAT	TGGGCAGATG	TGTGAGGCAC	CTGTGGTGAC	CCGAGAGTGG	GTGTTGGACA	5640
GTGTAGCACT	CTACCAGTGC	CAGGAGCTGG	ACACCTACCT	GATACCCCAG	ATCCCCCACA	5700
GCCACTACTG	A					5711

5,654,155

25

26

-continued

(2) INFORMATION FOR SEQ ID NO:2:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1863 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: Not Relevant
 - (D) TOPOLOGY: Not Relevant
- (i i) MOLECULE TYPE: protein
- (v i) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo sapiens
 - (B) STRAIN: BRCA1
- (v i i i) POSITION IN GENOME:
 - (A) CHROMOSOME/SEGMENT: 17
 - (B) MAP POSITION: 17q21

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met Asp Leu Ser Ala Leu Arg Val Glu Glu Val Gln Asn Val Ile Asn
1 5 10 15
Ala Met Gln Lys Ile Leu Glu Cys Pro Ile Cys Leu Glu Leu Ile Lys
20 25 30
Glu Pro Val Ser Thr Lys Cys Asp His Ile Phe Cys Lys Phe Cys Met
35 40 45
Leu Lys Leu Leu Asn Gln Lys Lys Gly Pro Ser Gln Cys Pro Leu Cys
50 55 60
Lys Asn Asp Ile Thr Lys Arg Ser Leu Gln Glu Ser Thr Arg Phe Ser
65 70 75 80
Gln Leu Val Glu Glu Leu Leu Lys Ile Ile Cys Ala Phe Gln Leu Asp
85 90 95
Thr Gly Leu Glu Tyr Ala Asn Ser Tyr Asn Phe Ala Lys Lys Glu Asn
100 105 110
Asn Ser Pro Glu His Leu Lys Asp Glu Val Ser Ile Ile Gln Ser Met
115 120 125
Gly Tyr Arg Asn Arg Ala Lys Arg Leu Leu Gln Ser Glu Pro Glu Asn
130 135 140
Pro Ser Leu Gln Glu Thr Ser Leu Ser Val Gln Leu Ser Asn Leu Gly
145 150 155 160
Thr Val Arg Thr Leu Arg Thr Lys Gln Arg Ile Gln Pro Gln Lys Thr
165 170 175
Ser Val Tyr Ile Glu Leu Gly Ser Asp Ser Ser Glu Asp Thr Val Asn
180 185 190
Lys Ala Thr Tyr Cys Ser Val Gly Asp Gln Glu Leu Leu Gln Ile Thr
195 200 205
Pro Gln Gly Thr Arg Asp Glu Ile Ser Leu Asp Ser Ala Lys Lys Ala
210 215 220
Ala Cys Glu Phe Ser Glu Thr Asp Val Thr Asn Thr Glu His His Gln
225 230 235 240
Pro Ser Asn Asn Asp Leu Asn Thr Thr Glu Lys Arg Ala Ala Glu Arg
245 250 255
His Pro Glu Lys Tyr Gln Gly Ser Ser Val Ser Asn Leu His Val Glu
260 265 270
Pro Cys Gly Thr Asn Thr His Ala Ser Ser Leu Gln His Glu Asn Ser
275 280 285
Ser Leu Leu Leu Thr Lys Asp Arg Met Asn Val Glu Lys Ala Glu Phe
290 295 300
Cys Asn Lys Ser Lys Gln Pro Gly Leu Ala Arg Ser Gln His Asn Arg
305 310 315 320

5,654,155

27

28

-continued

Trp	Ala	Gly	Ser	Lys	Glu	Thr	Cys	Asn	Asp	Arg	Arg	Thr	Pro	Ser	Thr
				325					330					335	
Glu	Lys	Lys	Val	Asp	Leu	Asn	Ala	Asp	Pro	Leu	Cys	Glu	Arg	Lys	Glu
			340					345					350		
Trp	Asn	Lys	Gln	Lys	Leu	Pro	Cys	Ser	Glu	Asn	Pro	Arg	Asp	Thr	Glu
		355					360					365			
Asp	Val	Pro	Trp	Ile	Thr	Leu	Asn	Ser	Ser	Ile	Gln	Lys	Val	Asn	Glu
	370					375					380				
Trp	Phe	Ser	Arg	Ser	Asp	Glu	Leu	Leu	Gly	Ser	Asp	Asp	Ser	His	Asp
	385				390					395					400
Gly	Glu	Ser	Glu	Ser	Asn	Ala	Lys	Val	Ala	Asp	Val	Leu	Asp	Val	Leu
				405					410					415	
Asn	Glu	Val	Asp	Glu	Tyr	Ser	Gly	Ser	Ser	Glu	Lys	Ile	Asp	Leu	Leu
			420					425					430		
Ala	Ser	Asp	Pro	His	Glu	Ala	Leu	Ile	Cys	Lys	Ser	Glu	Arg	Val	His
		435					440					445			
Ser	Lys	Ser	Val	Glu	Ser	Asn	Ile	Glu	Asp	Lys	Ile	Phe	Gly	Lys	Thr
	450					455					460				
Tyr	Arg	Lys	Lys	Ala	Ser	Leu	Pro	Asn	Leu	Ser	His	Val	Thr	Glu	Asn
	465				470					475					480
Leu	Ile	Ile	Gly	Ala	Phe	Val	Thr	Glu	Pro	Gln	Ile	Ile	Gln	Glu	Arg
				485					490					495	
Pro	Leu	Thr	Asn	Lys	Leu	Lys	Arg	Lys	Arg	Arg	Pro	Thr	Ser	Gly	Leu
			500					505					510		
His	Pro	Glu	Asp	Phe	Ile	Lys	Lys	Ala	Asp	Leu	Ala	Val	Gln	Lys	Thr
		515					520					525			
Pro	Glu	Met	Ile	Asn	Gln	Gly	Thr	Asn	Gln	Thr	Glu	Gln	Asn	Gly	Gln
	530					535					540				
Val	Met	Asn	Ile	Thr	Asn	Ser	Gly	His	Glu	Asn	Lys	Thr	Lys	Gly	Asp
	545				550				555						560
Ser	Ile	Gln	Asn	Glu	Lys	Asn	Pro	Asn	Pro	Ile	Glu	Ser	Leu	Glu	Lys
				565					570					575	
Glu	Ser	Ala	Phe	Lys	Thr	Lys	Ala	Glu	Pro	Ile	Ser	Ser	Ser	Ile	Ser
			580					585					590		
Asn	Met	Glu	Leu	Glu	Leu	Asn	Ile	His	Asn	Ser	Lys	Ala	Pro	Lys	Lys
		595					600					605			
Asn	Arg	Leu	Arg	Arg	Lys	Ser	Ser	Thr	Arg	His	Ile	His	Ala	Leu	Glu
	610					615					620				
Leu	Val	Val	Ser	Arg	Asn	Leu	Ser	Pro	Pro	Asn	Cys	Thr	Glu	Leu	Gln
	625				630					635					640
Ile	Asp	Ser	Cys	Ser	Ser	Ser	Glu	Glu	Ile	Lys	Lys	Lys	Lys	Tyr	Asn
				645					650					655	
Gln	Met	Pro	Val	Arg	His	Ser	Arg	Asn	Leu	Gln	Leu	Met	Glu	Gly	Lys
			660					665					670		
Glu	Pro	Ala	Thr	Gly	Ala	Lys	Lys	Ser	Asn	Lys	Pro	Asn	Glu	Gln	Thr
		675					680					685			
Ser	Lys	Arg	His	Asp	Ser	Asp	Thr	Phe	Pro	Glu	Leu	Lys	Leu	Thr	Asn
	690					695					700				
Ala	Pro	Gly	Ser	Phe	Thr	Lys	Cys	Ser	Asn	Thr	Ser	Glu	Leu	Lys	Glu
	705				710					715					720
Phe	Val	Asn	Pro	Ser	Leu	Pro	Arg	Glu	Glu	Lys	Glu	Glu	Lys	Leu	Glu
				725					730					735	
Thr	Val	Lys	Val	Ser	Asn	Asn	Ala	Glu	Asp	Pro	Lys	Asp	Leu	Met	Leu
			740					745					750		

A000121

5,654,155

29

30

-continued

Ser	Gly	Glu	Arg	Val	Leu	Gln	Thr	Glu	Arg	Ser	Val	Glu	Ser	Ser	Ser
Ile	Ser	Leu	Val	Pro	Gly	Thr	Asp	Tyr	Gly	Thr	Gln	Glu	Ser	Ile	Ser
Leu	Leu	Glu	Val	Ser	Thr	Leu	Gly	Lys	Ala	Lys	Thr	Glu	Pro	Asn	Lys
Cys	Val	Ser	Gln	Cys	Ala	Ala	Phe	Glu	Asn	Pro	Lys	Gly	Leu	Ile	His
Gly	Cys	Ser	Lys	Asp	Asn	Arg	Asn	Asp	Thr	Glu	Gly	Phe	Lys	Tyr	Pro
Leu	Gly	His	Glu	Val	Asn	His	Ser	Arg	Glu	Thr	Ser	Ile	Glu	Met	Glu
Glu	Ser	Glu	Leu	Asp	Ala	Gln	Tyr	Leu	Gln	Asn	Thr	Phe	Lys	Val	Ser
Lys	Arg	Gln	Ser	Phe	Ala	Leu	Phe	Ser	Asn	Pro	Gly	Asn	Ala	Glu	Glu
Glu	Cys	Ala	Thr	Phe	Ser	Ala	His	Ser	Gly	Ser	Leu	Lys	Lys	Gln	Ser
Pro	Lys	Val	Thr	Phe	Glu	Cys	Glu	Gln	Lys	Glu	Glu	Asn	Gln	Gly	Lys
Asn	Glu	Ser	Asn	Ile	Lys	Pro	Val	Gln	Thr	Val	Asn	Ile	Thr	Ala	Gly
Phe	Pro	Val	Val	Gly	Gln	Lys	Asp	Lys	Pro	Val	Asp	Asn	Ala	Lys	Cys
Ser	Ile	Lys	Gly	Gly	Ser	Arg	Phe	Cys	Leu	Ser	Ser	Gln	Phe	Arg	Gly
Asn	Glu	Thr	Gly	Leu	Ile	Thr	Pro	Asn	Lys	His	Gly	Leu	Leu	Gln	Asn
Pro	Tyr	Arg	Ile	Pro	Pro	Leu	Phe	Pro	Ile	Lys	Ser	Phe	Val	Lys	Thr
Lys	Cys	Lys	Lys	Asn	Leu	Leu	Glu	Glu	Asn	Phe	Glu	Glu	His	Ser	Met
Ser	Pro	Glu	Arg	Glu	Met	Gly	Asn	Glu	Asn	Ile	Pro	Ser	Thr	Val	Ser
Thr	Ile	Ser	Arg	Asn	Asn	Ile	Arg	Glu	Asn	Val	Phe	Lys	Gly	Ala	Ser
Ser	Ser	Asn	Ile	Asn	Glu	Val	Gly	Ser	Ser	Thr	Asn	Glu	Val	Gly	Ser
Ser	Ile	Asn	Glu	Ile	Gly	Ser	Ser	Asp	Glu	Asn	Ile	Gln	Ala	Glu	Leu
Gly	Arg	Asn	Arg	Gly	Pro	Lys	Leu	Asn	Ala	Met	Leu	Arg	Leu	Gly	Val
Leu	Gln	Pro	Glu	Val	Tyr	Lys	Gln	Ser	Leu	Pro	Gly	Ser	Asn	Cys	Lys
His	Pro	Glu	Ile	Lys	Lys	Gln	Glu	Tyr	Glu	Glu	Val	Val	Gln	Thr	Val
Asn	Thr	Asp	Phe	Ser	Pro	Tyr	Leu	Ile	Ser	Asp	Asn	Leu	Glu	Gln	Pro
Met	Gly	Ser	Ser	His	Ala	Ser	Gln	Val	Cys	Ser	Glu	Thr	Pro	Asp	Asp
Leu	Leu	Asp	Asp	Gly	Glu	Ile	Lys	Glu	Asp	Thr	Ser	Phe	Ala	Glu	Asn
Asp	Ile	Lys	Glu	Ser	Ser	Ala	Val	Phe	Ser	Lys	Ser	Val	Gln	Arg	Gly

A000122

5,654,155

31

32

-continued

1170					1175					1180					
Glu 1185	Leu	Ser	Arg	Ser	Pro 1190	Ser	Pro	Phe	Thr	His 1195	Thr	His	Leu	Ala	Gln 1200
Gly	Tyr	Arg	Arg	Gly 1205	Ala	Lys	Lys	Leu	Glu 1210	Ser	Ser	Glu	Glu	Asn 1215	Leu
Ser	Ser	Glu	Asp 1220	Glu	Glu	Leu	Pro	Cys 1225	Phe	Gln	His	Leu	Leu	Phe	Gly
Lys	Val	Asn 1235	Asn	Ile	Pro	Ser	Gln 1240	Ser	Thr	Arg	His	Ser 1245	Thr	Val	Ala
Thr	Glu 1250	Cys	Leu	Ser	Lys	Asn 1255	Thr	Glu	Glu	Asn 1260	Leu	Leu	Ser	Leu	Lys
Asn 1265	Ser	Leu	Asn	Asp	Cys 1270	Ser	Asn	Gln	Val	Ile 1275	Leu	Ala	Lys	Ala	Ser 1280
Gln	Glu	His	His	Leu 1285	Ser	Glu	Glu	Thr	Lys 1290	Cys	Ser	Ala	Ser	Leu	Phe 1295
Ser	Ser	Gln 1300	Cys	Ser	Glu	Leu	Glu	Asp 1305	Leu	Thr	Ala	Asn 1310	Thr	Asn	Thr
Gln	Asp	Pro 1315	Phe	Leu	Ile	Gly	Ser 1320	Ser	Lys	Gln	Met	Arg 1325	His	Gln	Ser
Glu 1330	Ser	Gln	Gly	Val	Gly	Leu 1335	Ser	Asp	Lys	Glu	Leu 1340	Val	Ser	Asp	Asp
Glu 1345	Glu	Arg	Gly	Thr	Gly 1350	Leu	Glu	Glu	Asn 1355	Asn	Gln	Glu	Glu	Gln	Ser 1360
Met	Asp	Ser	Asn 1365	Leu	Gly	Glu	Ala	Ala	Ser 1370	Gly	Cys	Glu	Ser	Glu	Thr 1375
Ser	Val	Ser	Glu 1380	Asp	Cys	Ser	Gly	Leu 1385	Ser	Ser	Gln	Ser	Asp 1390	Ile	Leu
Thr	Thr	Gln 1395	Gln	Arg	Asp	Thr	Met 1400	Gln	His	Asn	Leu 1405	Ile	Lys	Leu	Gln
Gln 1410	Glu	Met	Ala	Glu	Leu	Glu 1415	Ala	Val	Leu	Glu	Gln 1420	His	Gly	Ser	Gln
Pro 1425	Ser	Asn	Ser	Tyr	Pro 1430	Ser	Ile	Ile	Ser	Asp 1435	Ser	Ser	Ala	Leu	Glu 1440
Asp	Leu	Arg	Asn 1445	Pro	Glu	Gln	Ser	Thr	Ser 1450	Glu	Lys	Ala	Val	Leu	Thr 1455
Ser	Gln	Lys	Ser 1460	Ser	Glu	Tyr	Pro	Ile 1465	Ser	Gln	Asn	Pro	Glu 1470	Gly	Leu
Ser	Ala	Asp 1475	Lys	Phe	Glu	Val	Ser 1480	Ala	Asp	Ser	Ser	Thr 1485	Ser	Lys	Asn
Lys 1490	Glu	Pro	Gly	Val	Glu	Arg 1495	Ser	Ser	Pro	Ser	Lys 1500	Cys	Pro	Ser	Leu
Asp 1505	Asp	Arg	Trp	Tyr	Met 1510	His	Ser	Cys	Ser	Gly 1515	Ser	Leu	Gln	Asn	Arg 1520
Asn	Tyr	Pro	Ser	Gln 1525	Glu	Glu	Leu	Ile	Lys 1530	Val	Val	Asp	Val	Glu	Glu 1535
Gln	Gln	Leu	Glu 1540	Glu	Ser	Gly	Pro	His 1545	Asp	Leu	Thr	Glu 1550	Thr	Ser	Tyr
Leu	Pro	Arg 1555	Gln	Asp	Leu	Glu	Gly 1560	Thr	Pro	Tyr	Leu 1565	Glu	Ser	Gly	Ile
Ser	Leu 1570	Phe	Ser	Asp	Asp	Pro 1575	Glu	Ser	Asp	Pro	Ser 1580	Glu	Asp	Arg	Ala
Pro 1585	Glu	Ser	Ala	Arg	Val 1590	Gly	Asn	Ile	Pro	Ser 1595	Ser	Thr	Ser	Ala	Leu 1600

A000123

5,654,155

33

34

-continued

Lys	Val	Pro	Gln	Leu	Lys	Val	Ala	Glu	Ser	Ala	Gln	Gly	Pro	Ala	Ala
				1605					1610					1615	
Ala	His	Thr	Thr	Asp	Thr	Ala	Gly	Tyr	Asn	Ala	Met	Glu	Glu	Ser	Val
			1620					1625					1630		
Ser	Arg	Glu	Lys	Pro	Glu	Leu	Thr	Ala	Ser	Thr	Glu	Arg	Val	Asn	Lys
		1635					1640					1645			
Arg	Met	Ser	Met	Val	Val	Ser	Gly	Leu	Thr	Pro	Glu	Glu	Phe	Met	Leu
	1650					1655					1660				
Val	Tyr	Lys	Phe	Ala	Arg	Lys	His	His	Ile	Thr	Leu	Thr	Asn	Leu	Ile
1665					1670					1675					1680
Thr	Glu	Glu	Thr	Thr	His	Val	Val	Met	Lys	Thr	Asp	Ala	Glu	Phe	Val
			1685						1690					1695	
Cys	Glu	Arg	Thr	Leu	Lys	Tyr	Phe	Leu	Gly	Ile	Ala	Gly	Gly	Lys	Trp
			1700					1705					1710		
Val	Val	Ser	Tyr	Phe	Trp	Val	Thr	Gln	Ser	Ile	Lys	Glu	Arg	Lys	Met
		1715					1720					1725			
Leu	Asn	Glu	His	Asp	Phe	Glu	Val	Arg	Gly	Asp	Val	Val	Asn	Gly	Arg
	1730					1735					1740				
Asn	His	Gln	Gly	Pro	Lys	Arg	Ala	Arg	Glu	Ser	Gln	Asp	Arg	Lys	Ile
1745					1750					1755					1760
Phe	Arg	Gly	Leu	Glu	Ile	Cys	Cys	Tyr	Gly	Pro	Phe	Thr	Asn	Met	Pro
			1765						1770					1775	
Thr	Asp	Gln	Leu	Glu	Trp	Met	Val	Gln	Leu	Cys	Gly	Ala	Ser	Val	Val
			1780					1785					1790		
Lys	Glu	Leu	Ser	Ser	Phe	Thr	Leu	Gly	Thr	Gly	Val	His	Pro	Ile	Val
	1795						1800					1805			
Val	Val	Gln	Pro	Asp	Ala	Trp	Thr	Glu	Asp	Asn	Gly	Phe	His	Ala	Ile
	1810					1815					1820				
Gly	Gln	Met	Cys	Glu	Ala	Pro	Val	Val	Thr	Arg	Glu	Trp	Val	Leu	Asp
1825					1830					1835					1840
Ser	Val	Ala	Leu	Tyr	Gln	Cys	Gln	Glu	Leu	Asp	Thr	Tyr	Leu	Ile	Pro
			1845						1850					1855	
Gln	Ile	Pro	His	Ser	His	Tyr									
			1860												

(2) INFORMATION FOR SEQ ID NO:3:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 24 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: Not Relevant
 - (D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

- (v i) ORIGINAL SOURCE:
 - (B) STRAIN: 2F primer

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:3:

GAAGTTGTCA TTTTATAAAC CTTT

24

(2) INFORMATION FOR SEQ ID NO:4:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 22 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: Not Relevant
 - (D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

5,654,155

35

36

-continued

(v i) ORIGINAL SOURCE:
(B) STRAIN: 2R primer

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:4:

TGTCCTTTTCT TCCCTAGTAT GT

22

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: Not Relevant
- (D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(v i) ORIGINAL SOURCE:
(B) STRAIN: 3F primer

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:5:

TCCTGACACA GCAGACATT A

21

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: Not Relevant
- (D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(v i) ORIGINAL SOURCE:
(B) STRAIN: 3R primer

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:6:

TTGGATTTTC GTTCTCACTT A

21

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: Not Relevant
- (D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(v i) ORIGINAL SOURCE:
(B) STRAIN: 5F primer

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:7:

CTCTTAAGGG CAGTTGTGAG

20

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: Not Relevant
- (D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(v i) ORIGINAL SOURCE:
(B) STRAIN: 5R-M13* primer

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:8:

TTCCTACTGT GGTGCTTCC

20

5,654,155

37

38

-continued

(2) INFORMATION FOR SEQ ID NO:9:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 23 base pairs	
(B) TYPE: nucleic acid	
(C) STRANDEDNESS: Not Relevant	
(D) TOPOLOGY: linear	
(i i) MOLECULE TYPE: DNA (genomic)	
(v i) ORIGINAL SOURCE:	
(B) STRAIN: 67F primer	
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:9:	
CTTATTTTAG TGTCTTAAA AGG	2 3
(2) INFORMATION FOR SEQ ID NO:10:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 22 base pairs	
(B) TYPE: nucleic acid	
(C) STRANDEDNESS: Not Relevant	
(D) TOPOLOGY: linear	
(i i) MOLECULE TYPE: DNA (genomic)	
(v i) ORIGINAL SOURCE:	
(B) STRAIN: 6R	
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:10:	
TTTCATGGAC AGCACTTGAG TG	2 2
(2) INFORMATION FOR SEQ ID NO:11:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 23 base pairs	
(B) TYPE: nucleic acid	
(C) STRANDEDNESS: Not Relevant	
(D) TOPOLOGY: linear	
(i i) MOLECULE TYPE: DNA (genomic)	
(v i) ORIGINAL SOURCE:	
(B) STRAIN: 7F primer	
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:11:	
CACAACAAAG AGCATACATA GGG	2 3
(2) INFORMATION FOR SEQ ID NO:12:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 20 base pairs	
(B) TYPE: nucleic acid	
(C) STRANDEDNESS: Not Relevant	
(D) TOPOLOGY: linear	
(i i) MOLECULE TYPE: DNA (genomic)	
(v i) ORIGINAL SOURCE:	
(B) STRAIN: 67R primer	
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:12:	
TCTGGTTCAC TCTGTAGAAG	2 0
(2) INFORMATION FOR SEQ ID NO:13:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 21 base pairs	
(B) TYPE: nucleic acid	
(C) STRANDEDNESS: Not Relevant	
(D) TOPOLOGY: linear	

<div>(i i) MOLECULE TYPE: DNA (genomic)</div> <div>(v i) ORIGINAL SOURCE:<div>(B) STRAIN: 8F1 primer</div></div> <div>(x i) SEQUENCE DESCRIPTION: SEQ ID NO:13:</div> <div>TTCTCTTCAG GAGGAAAAGC A</div>		2 1
<div>(2) INFORMATION FOR SEQ ID NO:14:</div> <div>(i) SEQUENCE CHARACTERISTICS:<div>(A) LENGTH: 21 base pairs</div><div>(B) TYPE: nucleic acid</div><div>(C) STRANDEDNESS: Not Relevant</div><div>(D) TOPOLOGY: linear</div></div> <div>(i i) MOLECULE TYPE: DNA (genomic)</div> <div>(v i) ORIGINAL SOURCE:<div>(B) STRAIN: 8R1 primer</div></div> <div>(x i) SEQUENCE DESCRIPTION: SEQ ID NO:14:</div> <div>GCTGCCTACC ACAAATACAA A</div>		2 1
<div>(2) INFORMATION FOR SEQ ID NO:15:</div> <div>(i) SEQUENCE CHARACTERISTICS:<div>(A) LENGTH: 23 base pairs</div><div>(B) TYPE: nucleic acid</div><div>(C) STRANDEDNESS: Not Relevant</div><div>(D) TOPOLOGY: linear</div></div> <div>(i i) MOLECULE TYPE: DNA (genomic)</div> <div>(v i) ORIGINAL SOURCE:<div>(B) STRAIN: 9F primer</div></div> <div>(x i) SEQUENCE DESCRIPTION: SEQ ID NO:15:</div> <div>CCACAGTAGA TGCTCAGTAA ATA</div>		2 3
<div>(2) INFORMATION FOR SEQ ID NO:16:</div> <div>(i) SEQUENCE CHARACTERISTICS:<div>(A) LENGTH: 23 base pairs</div><div>(B) TYPE: nucleic acid</div><div>(C) STRANDEDNESS: Not Relevant</div><div>(D) TOPOLOGY: linear</div></div> <div>(i i) MOLECULE TYPE: DNA (genomic)</div> <div>(v i) ORIGINAL SOURCE:<div>(B) STRAIN: 9R primer</div></div> <div>(x i) SEQUENCE DESCRIPTION: SEQ ID NO:16:</div> <div>TAGGAAAATA CCAGCTTCAT AGA</div>		2 3
<div>(2) INFORMATION FOR SEQ ID NO:17:</div> <div>(i) SEQUENCE CHARACTERISTICS:<div>(A) LENGTH: 20 base pairs</div><div>(B) TYPE: nucleic acid</div><div>(C) STRANDEDNESS: Not Relevant</div><div>(D) TOPOLOGY: linear</div></div> <div>(i i) MOLECULE TYPE: DNA (genomic)</div> <div>(v i) ORIGINAL SOURCE:<div>(B) STRAIN: 10F primer</div></div> <div>(x i) SEQUENCE DESCRIPTION: SEQ ID NO:17:</div> <div>TGGTCAGCTT TCTGTAATCG</div>		2 0

5,654,155

41

42

-continued

(2) INFORMATION FOR SEQ ID NO:18:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 24 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: Not Relevant
 - (D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(v i) ORIGINAL SOURCE:

- (B) STRAIN: 10R primer

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:18:

GTATCTACCC ACTCTCTTCT TCAG 2 4

(2) INFORMATION FOR SEQ ID NO:19:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 19 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: Not Relevant
 - (D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(v i) ORIGINAL SOURCE:

- (B) STRAIN: 11AF primer

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:19:

CCACCTCCAA GGTGTATCA 1 9

(2) INFORMATION FOR SEQ ID NO:20:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: Not Relevant
 - (D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(v i) ORIGINAL SOURCE:

- (B) STRAIN: 11AR primer

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:20:

TGTTATGTTG GCTCCTTGCT 2 0

(2) INFORMATION FOR SEQ ID NO:21:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 22 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: Not Relevant
 - (D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(v i) ORIGINAL SOURCE:

- (B) STRAIN: 11BF1 primer

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:21:

CACTAAAGAC AGAATGAATC TA 2 2

(2) INFORMATION FOR SEQ ID NO:22:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 22 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: Not Relevant

5,654,155

43

44

-continued

(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(v i) ORIGINAL SOURCE:
(B) STRAIN: 11BR1 primer

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:22:

GAAGAACCAG AATATTCATC TA 2 2

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: Not Relevant
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(v i) ORIGINAL SOURCE:
(B) STRAIN: 11CF1 primer

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:23:

TGATGGGGAG TCTGAATCAA 2 0

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: Not Relevant
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(v i) ORIGINAL SOURCE:
(B) STRAIN: 11CR1 primer

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:24:

TCTGCTTTCT TGATAAAATC CT 2 2

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: Not Relevant
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(v i) ORIGINAL SOURCE:
(B) STRAIN: 11DF1 primer

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:25:

AGCGTCCCCT CACAAATAAA 2 0

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: Not Relevant
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(v i) ORIGINAL SOURCE:
(B) STRAIN: 11DR1 primer

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:26:

5,654,155

45

46

-continued

TCAAGCGCAT GAATATGCCT	2 0
(2) INFORMATION FOR SEQ ID NO:27:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 22 base pairs	
(B) TYPE: nucleic acid	
(C) STRANDEDNESS: Not Relevant	
(D) TOPOLOGY: linear	
(i i) MOLECULE TYPE: DNA (genomic)	
(v i) ORIGINAL SOURCE:	
(B) STRAIN: 11EF primer	
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:27:	
GTATAAGCAA TATGGAAGCTC GA	2 2
(2) INFORMATION FOR SEQ ID NO:28:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 23 base pairs	
(B) TYPE: nucleic acid	
(C) STRANDEDNESS: Not Relevant	
(D) TOPOLOGY: linear	
(i i) MOLECULE TYPE: DNA (genomic)	
(v i) ORIGINAL SOURCE:	
(B) STRAIN: 11ER primer	
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:28:	
TTAAGTTCACT GGTATTTGAA CA	2 3
(2) INFORMATION FOR SEQ ID NO:29:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 20 base pairs	
(B) TYPE: nucleic acid	
(C) STRANDEDNESS: Not Relevant	
(D) TOPOLOGY: linear	
(i i) MOLECULE TYPE: DNA (genomic)	
(v i) ORIGINAL SOURCE:	
(B) STRAIN: 11FF primer	
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:29:	
GACAGCGATA CTTTCCCAGA	2 0
(2) INFORMATION FOR SEQ ID NO:30:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 21 base pairs	
(B) TYPE: nucleic acid	
(C) STRANDEDNESS: Not Relevant	
(D) TOPOLOGY: linear	
(i i) MOLECULE TYPE: DNA (genomic)	
(v i) ORIGINAL SOURCE:	
(B) STRAIN: 11FR primer	
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:30:	
TGGAACAACC ATGAATTAGT C	2 1
(2) INFORMATION FOR SEQ ID NO:31:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 20 base pairs	

5,654,155

47

48

-continued

(B) TYPE: nucleic acid
(C) STRANDEDNESS: Not Relevant
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(v i) ORIGINAL SOURCE:
(B) STRAIN: 11GF primer

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:31:

GGAAGTTAGC ACTCTAGGGA

20

(2) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: Not Relevant
- (D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(v i) ORIGINAL SOURCE:
(B) STRAIN: 11GR primer

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:32:

GCAGTGATAT TAACTGTCTG TA

22

(2) INFORMATION FOR SEQ ID NO:33:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: Not Relevant
- (D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(v i) ORIGINAL SOURCE:
(B) STRAIN: 11HF primer

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:33:

TGGGTCCTTA AAGAAACAAA GT

22

(2) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: Not Relevant
- (D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(v i) ORIGINAL SOURCE:
(B) STRAIN: 11HR primer

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:34:

TCAGGTGACA TTGAATCTTC C

21

(2) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: Not Relevant
- (D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(v i) ORIGINAL SOURCE:
(B) STRAIN: 11IF primer

5,654,155

49

50

-continued

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:35:	
CCACTTTTTC CCATCAAGTC A	2 1
(2) INFORMATION FOR SEQ ID NO:36:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 22 base pairs	
(B) TYPE: nucleic acid	
(C) STRANDEDNESS: Not Relevant	
(D) TOPOLOGY: linear	
(i i) MOLECULE TYPE: DNA (genomic)	
(v i) ORIGINAL SOURCE:	
(B) STRAIN: 11IR primer	
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:36:	
TCAGGATGCT TACAATTACT TC	2 2
(2) INFORMATION FOR SEQ ID NO:37:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 23 base pairs	
(B) TYPE: nucleic acid	
(C) STRANDEDNESS: Not Relevant	
(D) TOPOLOGY: linear	
(i i) MOLECULE TYPE: DNA (genomic)	
(v i) ORIGINAL SOURCE:	
(B) STRAIN: 11JF primer	
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:37:	
CAAAATTGAA TGCTATGCTT AGA	2 3
(2) INFORMATION FOR SEQ ID NO:38:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 20 base pairs	
(B) TYPE: nucleic acid	
(C) STRANDEDNESS: Not Relevant	
(D) TOPOLOGY: linear	
(i i) MOLECULE TYPE: DNA (genomic)	
(v i) ORIGINAL SOURCE:	
(B) STRAIN: 11JR primer	
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:38:	
TCGGTAACCC TGAGCCAAAT	2 0
(2) INFORMATION FOR SEQ ID NO:39:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 20 base pairs	
(B) TYPE: nucleic acid	
(C) STRANDEDNESS: Not Relevant	
(D) TOPOLOGY: linear	
(i i) MOLECULE TYPE: DNA (genomic)	
(v i) ORIGINAL SOURCE:	
(B) STRAIN: 11KF primer	
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:39:	
GCAAAAGCGT CCAGAAAGGA	2 0
(2) INFORMATION FOR SEQ ID NO:40:	

5,654,155

51

52

-continued

<div>(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 22 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: Not Relevant (D) TOPOLOGY: linear</div>	
<div>(i i) MOLECULE TYPE: DNA (genomic)</div>	
<div>(v i) ORIGINAL SOURCE: (B) STRAIN: 11KR-1 primer</div>	
<div>(x i) SEQUENCE DESCRIPTION: SEQ ID NO:40:</div>	
T A T T T G C A G T C A A G T C T T C C A A	2 2
<div>(2) INFORMATION FOR SEQ ID NO:41:</div>	
<div>(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 21 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: Not Relevant (D) TOPOLOGY: linear</div>	
<div>(i i) MOLECULE TYPE: DNA (genomic)</div>	
<div>(v i) ORIGINAL SOURCE: (B) STRAIN: 11LF-1 primer</div>	
<div>(x i) SEQUENCE DESCRIPTION: SEQ ID NO:41:</div>	
G T A A T A T T G G C A A A G G C A T C T	2 1
<div>(2) INFORMATION FOR SEQ ID NO:42:</div>	
<div>(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 22 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: Not Relevant (D) TOPOLOGY: linear</div>	
<div>(i i) MOLECULE TYPE: DNA (genomic)</div>	
<div>(v i) ORIGINAL SOURCE: (B) STRAIN: 11LR primer</div>	
<div>(x i) SEQUENCE DESCRIPTION: SEQ ID NO:42:</div>	
T A A A A T G T G C T C C C C A A A A G C A	2 2
<div>(2) INFORMATION FOR SEQ ID NO:43:</div>	
<div>(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 20 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: Not Relevant (D) TOPOLOGY: linear</div>	
<div>(i i) MOLECULE TYPE: DNA (genomic)</div>	
<div>(v i) ORIGINAL SOURCE: (B) STRAIN: 12F primer</div>	
<div>(x i) SEQUENCE DESCRIPTION: SEQ ID NO:43:</div>	
G T C C T G C C A A T G A G A A G A A A	2 0
<div>(2) INFORMATION FOR SEQ ID NO:44:</div>	
<div>(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 21 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: Not Relevant (D) TOPOLOGY: linear</div>	
<div>(i i) MOLECULE TYPE: DNA (genomic)</div>	

5,654,155

53

54

-continued

<div>(v i) ORIGINAL SOURCE: (B) STRAIN: 12R primer</div>	
<div>(x i) SEQUENCE DESCRIPTION: SEQ ID NO:44:</div>	
T G T C A G C A A A C C T A A G A A T G T	2 1
<div>(2) INFORMATION FOR SEQ ID NO:45:</div>	
<div>(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 21 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: Not Relevant (D) TOPOLOGY: linear</div>	
<div>(i i) MOLECULE TYPE: DNA (genomic)</div>	
<div>(v i) ORIGINAL SOURCE: (B) STRAIN: 13F primer</div>	
<div>(x i) SEQUENCE DESCRIPTION: SEQ ID NO:45:</div>	
A A T G G A A A G C T T C T C A A A G T A	2 1
<div>(2) INFORMATION FOR SEQ ID NO:46:</div>	
<div>(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 21 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: Not Relevant (D) TOPOLOGY: linear</div>	
<div>(i i) MOLECULE TYPE: DNA (genomic)</div>	
<div>(v i) ORIGINAL SOURCE: (B) STRAIN: 13R primer</div>	
<div>(x i) SEQUENCE DESCRIPTION: SEQ ID NO:46:</div>	
A T G T T G G A G C T A G G T C C T T A C	2 1
<div>(2) INFORMATION FOR SEQ ID NO:47:</div>	
<div>(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 22 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: Not Relevant (D) TOPOLOGY: linear</div>	
<div>(i i) MOLECULE TYPE: DNA (genomic)</div>	
<div>(v i) ORIGINAL SOURCE: (B) STRAIN: 14F primer</div>	
<div>(x i) SEQUENCE DESCRIPTION: SEQ ID NO:47:</div>	
C T A A C C T G A A T T A T C A C T A T C A	2 2
<div>(2) INFORMATION FOR SEQ ID NO:48:</div>	
<div>(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 21 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: Not Relevant (D) TOPOLOGY: linear</div>	
<div>(i i) MOLECULE TYPE: DNA (genomic)</div>	
<div>(v i) ORIGINAL SOURCE: (B) STRAIN: 14R primer</div>	
<div>(x i) SEQUENCE DESCRIPTION: SEQ ID NO:48:</div>	
G T G T A T A A A T G C C T G T A T G C A	2 1

5,654,155

55

56

-continued

(2) INFORMATION FOR SEQ ID NO:49:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 19 base pairs	
(B) TYPE: nucleic acid	
(C) STRANDEDNESS: Not Relevant	
(D) TOPOLOGY: linear	
(i i) MOLECULE TYPE: DNA (genomic)	
(v i) ORIGINAL SOURCE:	
(B) STRAIN: 15F primer	
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:49:	
T G G C T G C C C A G G A A G T A T G	1 9
(2) INFORMATION FOR SEQ ID NO:50:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 23 base pairs	
(B) TYPE: nucleic acid	
(C) STRANDEDNESS: Not Relevant	
(D) TOPOLOGY: linear	
(i i) MOLECULE TYPE: DNA (genomic)	
(v i) ORIGINAL SOURCE:	
(B) STRAIN: 15R primer	
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:50:	
A A C C A G A A T A T C T T T A T G T A G G A	2 3
(2) INFORMATION FOR SEQ ID NO:51:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 22 base pairs	
(B) TYPE: nucleic acid	
(C) STRANDEDNESS: Not Relevant	
(D) TOPOLOGY: linear	
(i i) MOLECULE TYPE: DNA (genomic)	
(v i) ORIGINAL SOURCE:	
(B) STRAIN: 16F primer	
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:51:	
A A T T C T T A A C A G A G A C C A G A A C	2 2
(2) INFORMATION FOR SEQ ID NO:52:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 22 base pairs	
(B) TYPE: nucleic acid	
(C) STRANDEDNESS: Not Relevant	
(D) TOPOLOGY: linear	
(i i) MOLECULE TYPE: DNA (genomic)	
(v i) ORIGINAL SOURCE:	
(B) STRAIN: 16R primer	
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:52:	
A A A A C T C T T T C C A G A A T G T T G T	2 2
(2) INFORMATION FOR SEQ ID NO:53:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 20 base pairs	
(B) TYPE: nucleic acid	
(C) STRANDEDNESS: Not Relevant	
(D) TOPOLOGY: linear	

5,654,155

57

58

-continued

<div>(i i) MOLECULE TYPE: DNA (genomic)</div> <div>(v i) ORIGINAL SOURCE:<div>(B) STRAIN: 17F primer</div></div> <div>(x i) SEQUENCE DESCRIPTION: SEQ ID NO:53:</div> <div>GTGTAGAACG TGCAGGATTG</div>		20
<div>(2) INFORMATION FOR SEQ ID NO:54:</div> <div>(i) SEQUENCE CHARACTERISTICS:<div>(A) LENGTH: 18 base pairs<div>(B) TYPE: nucleic acid<div>(C) STRANDEDNESS: Not Relevant<div>(D) TOPOLOGY: linear</div></div></div></div></div> <div>(i i) MOLECULE TYPE: DNA (genomic)</div> <div>(v i) ORIGINAL SOURCE:<div>(B) STRAIN: 17R primer</div></div> <div>(x i) SEQUENCE DESCRIPTION: SEQ ID NO:54:</div> <div>TGCGCTCATG TGGTTTTA</div>		18
<div>(2) INFORMATION FOR SEQ ID NO:55:</div> <div>(i) SEQUENCE CHARACTERISTICS:<div>(A) LENGTH: 21 base pairs<div>(B) TYPE: nucleic acid<div>(C) STRANDEDNESS: Not Relevant<div>(D) TOPOLOGY: linear</div></div></div></div></div> <div>(i i) MOLECULE TYPE: DNA (genomic)</div> <div>(v i) ORIGINAL SOURCE:<div>(B) STRAIN: 18F primer</div></div> <div>(x i) SEQUENCE DESCRIPTION: SEQ ID NO:55:</div> <div>GGCTCTTTAG CTTCTTAGGA C</div>		21
<div>(2) INFORMATION FOR SEQ ID NO:56:</div> <div>(i) SEQUENCE CHARACTERISTICS:<div>(A) LENGTH: 20 base pairs<div>(B) TYPE: nucleic acid<div>(C) STRANDEDNESS: Not Relevant<div>(D) TOPOLOGY: linear</div></div></div></div></div> <div>(i i) MOLECULE TYPE: DNA (genomic)</div> <div>(v i) ORIGINAL SOURCE:<div>(B) STRAIN: 18R primer</div></div> <div>(x i) SEQUENCE DESCRIPTION: SEQ ID NO:56:</div> <div>GAGACCATTT TCCCAGCATC</div>		20
<div>(2) INFORMATION FOR SEQ ID NO:57:</div> <div>(i) SEQUENCE CHARACTERISTICS:<div>(A) LENGTH: 20 base pairs<div>(B) TYPE: nucleic acid<div>(C) STRANDEDNESS: Not Relevant<div>(D) TOPOLOGY: linear</div></div></div></div></div> <div>(i i) MOLECULE TYPE: DNA (genomic)</div> <div>(v i) ORIGINAL SOURCE:<div>(B) STRAIN: 19F primer</div></div> <div>(x i) SEQUENCE DESCRIPTION: SEQ ID NO:57:</div> <div>CTGTCATTCT TCCTGTGCTC</div>		20

5,654,155

59

60

-continued

(2) INFORMATION FOR SEQ ID NO:58:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 21 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: Not Relevant
 - (D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(v i) ORIGINAL SOURCE:

- (B) STRAIN: 19R primer

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:58:

C A T T G T T A A G G A A A G T G G T G C 2 1

(2) INFORMATION FOR SEQ ID NO:59:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: Not Relevant
 - (D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(v i) ORIGINAL SOURCE:

- (B) STRAIN: 20F primer

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:59:

A T A T G A C G T G T C T G C T C C A C 2 0

(2) INFORMATION FOR SEQ ID NO:60:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: Not Relevant
 - (D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(v i) ORIGINAL SOURCE:

- (B) STRAIN: 20R primer

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:60:

G G G A A T C C A A A T T A C A C A G C 2 0

(2) INFORMATION FOR SEQ ID NO:61:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 22 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: Not Relevant
 - (D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(v i) ORIGINAL SOURCE:

- (B) STRAIN: 21F primer

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:61:

A A G C T C T T C C T T T T G A A A G T C 2 2

(2) INFORMATION FOR SEQ ID NO:62:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 22 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: Not Relevant

5,654,155

61

62

-continued

(D) TOPOLOGY: linear	
(i i) MOLECULE TYPE: DNA (genomic)	
(v i) ORIGINAL SOURCE:	
(B) STRAIN: 21R primer	
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:62:	
GTAGAGAAAT AGAATAGCCT CT	2 2
(2) INFORMATION FOR SEQ ID NO:63:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 20 base pairs	
(B) TYPE: nucleic acid	
(C) STRANDEDNESS: Not Relevant	
(D) TOPOLOGY: linear	
(i i) MOLECULE TYPE: DNA (genomic)	
(v i) ORIGINAL SOURCE:	
(B) STRAIN: 22F primer	
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:63:	
TCCCATTGAG AGGTCTTGCT	2 0
(2) INFORMATION FOR SEQ ID NO:64:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 20 base pairs	
(B) TYPE: nucleic acid	
(C) STRANDEDNESS: Not Relevant	
(D) TOPOLOGY: linear	
(i i) MOLECULE TYPE: DNA (genomic)	
(v i) ORIGINAL SOURCE:	
(B) STRAIN: 22R primer	
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:64:	
GAGAAGACTT CTGAGGCTAC	2 0
(2) INFORMATION FOR SEQ ID NO:65:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 21 base pairs	
(B) TYPE: nucleic acid	
(C) STRANDEDNESS: Not Relevant	
(D) TOPOLOGY: linear	
(i i) MOLECULE TYPE: DNA (genomic)	
(v i) ORIGINAL SOURCE:	
(B) STRAIN: 23F-1 primer	
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:65:	
TGAAGTGACA GTTCCAGTAG T	2 1
(2) INFORMATION FOR SEQ ID NO:66:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 23 base pairs	
(B) TYPE: nucleic acid	
(C) STRANDEDNESS: Not Relevant	
(D) TOPOLOGY: linear	
(i i) MOLECULE TYPE: DNA (genomic)	
(v i) ORIGINAL SOURCE:	
(B) STRAIN: 23R-1 primer	
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:66:	

5,654,155

63

64

-continued

CATTTTAGCC ATTCATTCAA CAA 2 3

(2) INFORMATION FOR SEQ ID NO:67:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 22 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: Not Relevant
 - (D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(v i) ORIGINAL SOURCE:

- (B) STRAIN: 24F primer

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:67:

ATGAATTGAC ACTAATCTCT GC 2 2

(2) INFORMATION FOR SEQ ID NO:68:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 21 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: Not Relevant
 - (D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(v i) ORIGINAL SOURCE:

- (B) STRAIN: 24R primer

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:68:

GTAGCCAGGA CAGTAGAAGG A 2 1

(2) INFORMATION FOR SEQ ID NO:69:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: Not Relevant
 - (D) TOPOLOGY: Not Relevant

(i i) MOLECULE TYPE: DNA (genomic)

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:69:

GCAAAAGCGT CCAGAAAGGA 2 0

(2) INFORMATION FOR SEQ ID NO:70:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: Not Relevant
 - (D) TOPOLOGY: Not Relevant

(i i) MOLECULE TYPE: DNA (genomic)

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:70:

AGTCTTCAA TTTACTGCAC 2 0

(2) INFORMATION FOR SEQ ID NO:71:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 14 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: Not Relevant
 - (D) TOPOLOGY: Not Relevant

(i i) MOLECULE TYPE: DNA (genomic)

5,654,155

65

66

-continued

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:71:	
GAACACAGGA GAAT	1 4
(2) INFORMATION FOR SEQ ID NO:72:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 14 base pairs	
(B) TYPE: nucleic acid	
(C) STRANDEDNESS: Not Relevant	
(D) TOPOLOGY: Not Relevant	
(i i) MOLECULE TYPE: DNA (genomic)	
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:72:	
TAAGAACACA GGAG	1 4
(2) INFORMATION FOR SEQ ID NO:73:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 16 base pairs	
(B) TYPE: nucleic acid	
(C) STRANDEDNESS: Not Relevant	
(D) TOPOLOGY: Not Relevant	
(i i) MOLECULE TYPE: DNA (genomic)	
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:73:	
GAACACAGAG GAGAAT	1 6
(2) INFORMATION FOR SEQ ID NO:74:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 16 base pairs	
(B) TYPE: nucleic acid	
(C) STRANDEDNESS: Not Relevant	
(D) TOPOLOGY: Not Relevant	
(i i) MOLECULE TYPE: DNA (genomic)	
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:74:	
TAAGAACACA GAGGAG	1 6

- We claim:
1. An isolated consensus DNA sequence of the BRCA1 coding sequence as set forth in SEQ ID NO: 1.
 2. A method of identifying individuals having a BRCA1 gene with a BRCA1 coding sequence not associated with breast or ovarian cancer comprising:
 - a) amplifying a DNA fragment of an individual's BRCA1 coding sequence using an oligonucleotide primer which specifically hybridizes to sequences within the gene;
 - b) sequencing said amplified fragment by dideoxy sequencing;
 - c) repeating steps (a) and (b) until said individual's BRCA1 coding sequence is completely sequenced;
 - d) comparing the sequence of said amplified DNA to the sequence of SEQ. ID. NO: 1;
 - e) determining the presence or absence of each of the following polymorphic variations in said individual's BRCA1 coding sequence:
AGC and ACT at position 2201,
TTG and CTG at position 2430,
CCG and CTG at position 2731,
GAA and GGA at position 3232,
AAA and AGA at position 3667,
TCT and TCC at position 4427,
and ACT and GGT at position 4956;
 - f) determining any sequence differences between said individual's BRCA1 coding sequences and SEQ. ID. NO: 1 wherein the presence of any of the said polymorphic variations and the absence of a polymorphism outside of positions 2201, 2430, 2731, 3232, 3667, 4427, and 4956, is correlated with an absence of increased genetic susceptibility to breast or ovarian cancer resulting from a BRCA1 mutation in the BRCA1 coding sequence.
 3. A method according to claim 2 wherein said oligonucleotide primer is labeled with a radiolabel, a fluorescent label, a bioluminescent label, a chemiluminescent label or an enzyme label.
 4. A method of detecting an increased genetic susceptibility to breast and ovarian cancer in an individual resulting from the presence of a mutation in the BRCA1 coding sequence, comprising:
 - a) amplifying a DNA fragment of an individual's BRCA1 coding sequence using an oligonucleotide primer which specifically hybridizes to sequences within the gene;
 - b) sequencing said amplified fragment by dideoxy sequencing;

5,654,155

67

- c) repeating steps (a) and (b) until said individual's BRCA1 coding sequence is completely sequenced;
- d) comparing the sequence of said amplified DNA to the sequence of SEQ. ID. NO: 1;
- e) determining any sequence differences between said individual's BRCA1 coding sequences and SEQ. ID. NO: 1 to determine the presence or absence of polymorphisms in said individual's BRCA coding sequences wherein a polymorphism which is not any of the following:
AGC or AGT at position 2201,

68

TTG or CTG at position 2430,
CCG or CTG at position 2731,
GAA or GGA at position 3232,
AAA or AGA at position 3667,
TCT or TCC at position 4427,
and AGT or GGT at position 4956;
is correlated with the potential of increased genetic susceptibility to breast or ovarian cancer resulting from a BRCA1 mutation in the BRCA1 coding sequence.

* * * * *

UNITED STATES PATENT AND TRADEMARK OFFICE
CERTIFICATE OF CORRECTION

PATENT NO. : 5,654,155
DATED : August 5, 1997
INVENTOR(S) : Brenda S. Critz, et al

It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

Column 15, line 34, replace "bellow" with --below--;

Column 18, line 57, replace "Baudet" with --Beaudet--;

Column 20, line 5, replace "Baudet" with --Beaudet--;

Signed and Sealed this
Twenty-eighth Day of April, 1998



Attest:

BRUCE LEHMAN

Attesting Officer

Commissioner of Patents and Trademarks

UNITED STATES PATENT AND TRADEMARK OFFICE
CERTIFICATE OF CORRECTION

PATENT NO. : 5,654,155
DATED : August 5, 1997
INVENTOR(S) : Brenda S. Critz, et al

It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

Column 15, line 34, replace "bellow" with --below--;

Column 18, line 57, replace "Baudet" with --Beaudet--;

Column 20, line 5, replace "Baudet" with --Beaudet--;

Signed and Sealed this

Twenty-eighth Day of April, 1998



Attest:

BRUCE LEHMAN

Attesting Officer

Commissioner of Patents and Trademarks

UNITED STATES PATENT AND TRADEMARK OFFICE
CERTIFICATE OF CORRECTION

PATENT NO. : 5,654,155
APPLICATION NO. : 08/598591
DATED : August 5, 1997
INVENTOR(S) : Patricia D. Murphy et al.

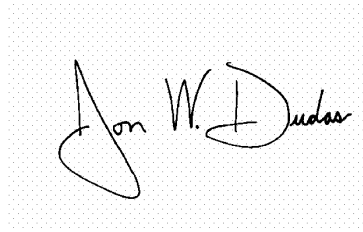
Page 1 of 1

It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

Column 66, line 44 (Claim 2) – please replace “ACT” with --AGT--

Signed and Sealed this

Thirtieth Day of January, 2007

A handwritten signature in black ink, reading "Jon W. Dudas", is written over a rectangular area with a light gray dotted background.

JON W. DUDAS
Director of the United States Patent and Trademark Office

A000144

UNITED STATES PATENT AND TRADEMARK OFFICE
CERTIFICATE OF CORRECTION

PATENT NO. : 5,654,155
APPLICATION NO. : 08/598591
DATED : August 5, 1997
INVENTOR(S) : Patricia D. Murphy et al.

Page 1 of 1

It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

Column 65, line 63 (Claim 2) - please replace "ACT" with --AGT--

Signed and Sealed this

Twenty-eighth Day of October, 2008

A handwritten signature in black ink, reading "Jon W. Dudas". The signature is stylized, with a large loop for the "J" and a cursive "Dudas".

JON W. DUDAS
Director of the United States Patent and Trademark Office

A000145



US005747282A

United States Patent

Skolnick et al.

[19]

[11] Patent Number:

[45] Date of Patent:

5,747,282

May 5, 1998

- [54] 17Q-LINKED BREAST AND OVARIAN
CANCER SUSCEPTIBILITY GENE
- [75] Inventors: Mark H. Skolnick; David E. Goldgar;
Yoshio Miki; Jeff Swenson; Alexander
Kamb; Keith D. Harshman; Donna
M. Shattuck-Eidens; Sean V.
Tavtigian, all of Salt Lake City, Utah;
Roger W. Wiseman; P. Andrew
Futreal, both of Durham, N.C.
- [73] Assignees: Myraid Genetics, Inc.; University of
Utah Research Foundation, both of
Salt Lake City, Utah; The United
States of America as represented by
the Secretary of Health and Human
Services, Washington, D.C.

- [21] Appl. No.: 483,554
- [22] Filed: Jun. 7, 1995

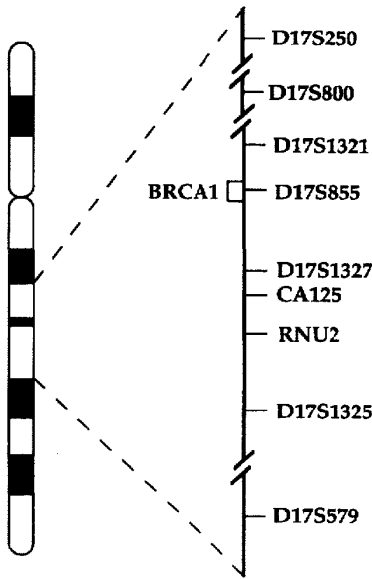
Related U.S. Application Data

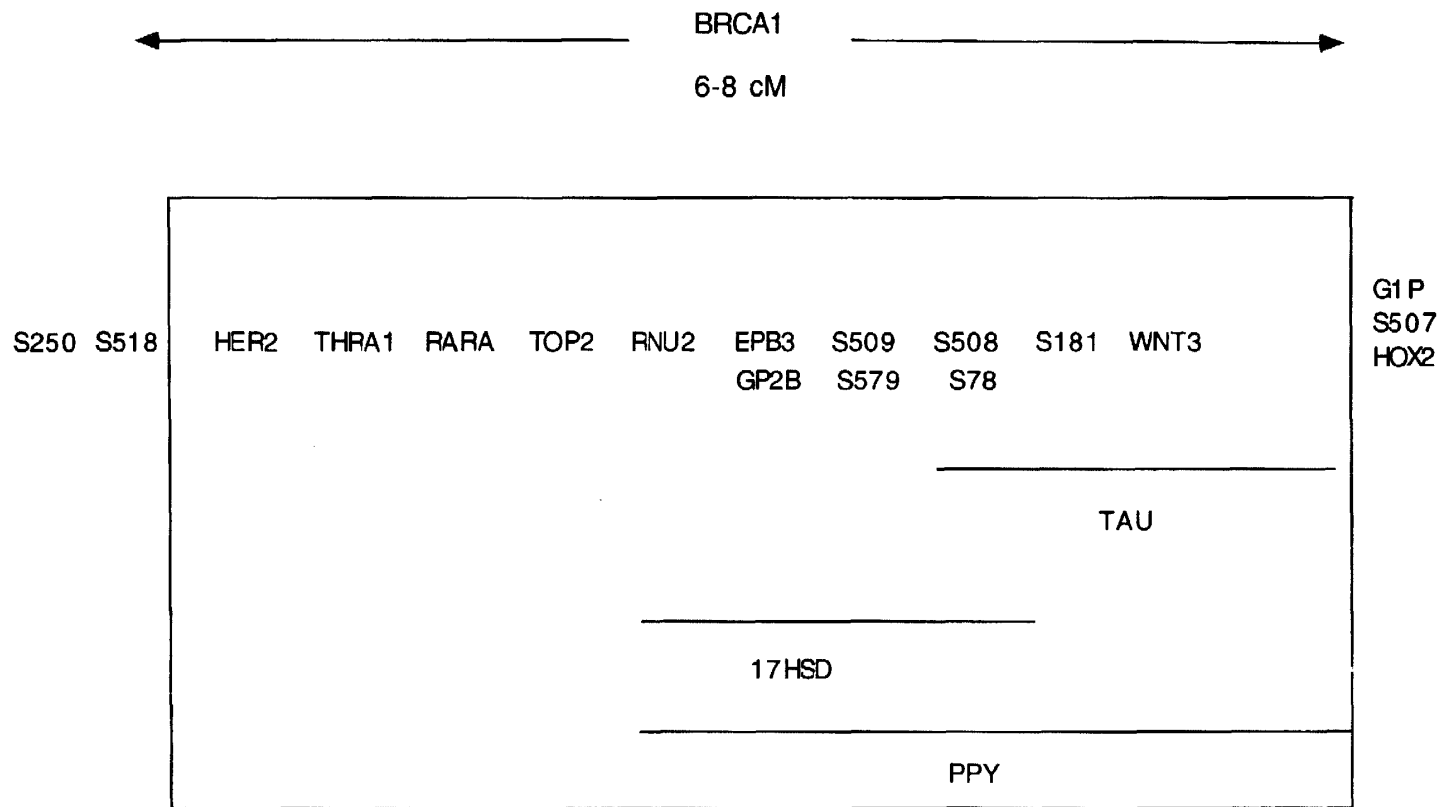
- [63] Continuation-in-part of Ser. No. 409,305, Mar. 24, 1995,
abandoned, which is a continuation-in-part of Ser. No.
348,824, Nov. 29, 1994, abandoned, which is a continuation-
in-part of Ser. No. 308,104, Sep. 16, 1994, which is a
continuation-in-part of Ser. No. 300,266, Sep. 2, 1994,
abandoned, which is a continuation-in-part of Ser. No.
289,221, Aug. 12, 1994, abandoned.
- [51] Int. Cl.⁶ C12P 21/06; C12Q 1/68;
C07H 21/04
- [52] U.S. Cl. 435/69.1; 435/6; 435/320.1;
435/325; 536/23.5; 536/24.31; 536/24.33;
935/9; 935/60; 935/66
- [58] Field of Search 536/23.1, 23.5,
536/24.33; 435/69.1, 6, 7.1, 91.1, 91.2,
375, 320.1; 514/44

- [56] References Cited
PUBLICATIONS
- E. Marshall (1995) Science 269:1050–1055.
Vila et al (1995) Targeted Gene Therapy 9:190–199.
Molecular Biology of the Gene, 4th Edition, vol. 1 Eds. J. D.
Watson et al., 1987, p. 313.
- Primary Examiner—Bruce R. Campell
Assistant Examiner—Abdur Razzaque
Attorney, Agent, or Firm—Venable, Baetjer, Howard &
Civiletti, LLP

- [57] ABSTRACT
- The present invention relates generally to the field of human
genetics. Specifically, the present invention relates to meth-
ods and materials used to isolate and detect a human breast
and ovarian cancer predisposing gene (BRCA1), some
mutant alleles of which cause susceptibility to cancer, in
particular breast and ovarian cancer. More specifically, the
invention relates to germline mutations in the BRCA1 gene
and their use in the diagnosis of predisposition to breast and
ovarian cancer. The present invention further relates to
somatic mutations in the BRCA1 gene in human breast and
ovarian cancer and their use in the diagnosis and prognosis
of human breast and ovarian cancer. Additionally, the inven-
tion relates to somatic mutations in the BRCA1 gene in other
human cancers and their use in the diagnosis and prognosis
of human cancers. The invention also relates to the therapy
of human cancers which have a mutation in the BRCA1
gene, including gene therapy, protein replacement therapy
and protein mimetics. The invention further relates to the
screening of drugs for cancer therapy. Finally, the invention
relates to the screening of the BRCA1 gene for mutations,
which are useful for diagnosing the predisposition to breast
and ovarian cancer.

20 Claims, 18 Drawing Sheets





Map of the early onset breast and ovarian cancer region (BRCA1)

FIG. 1

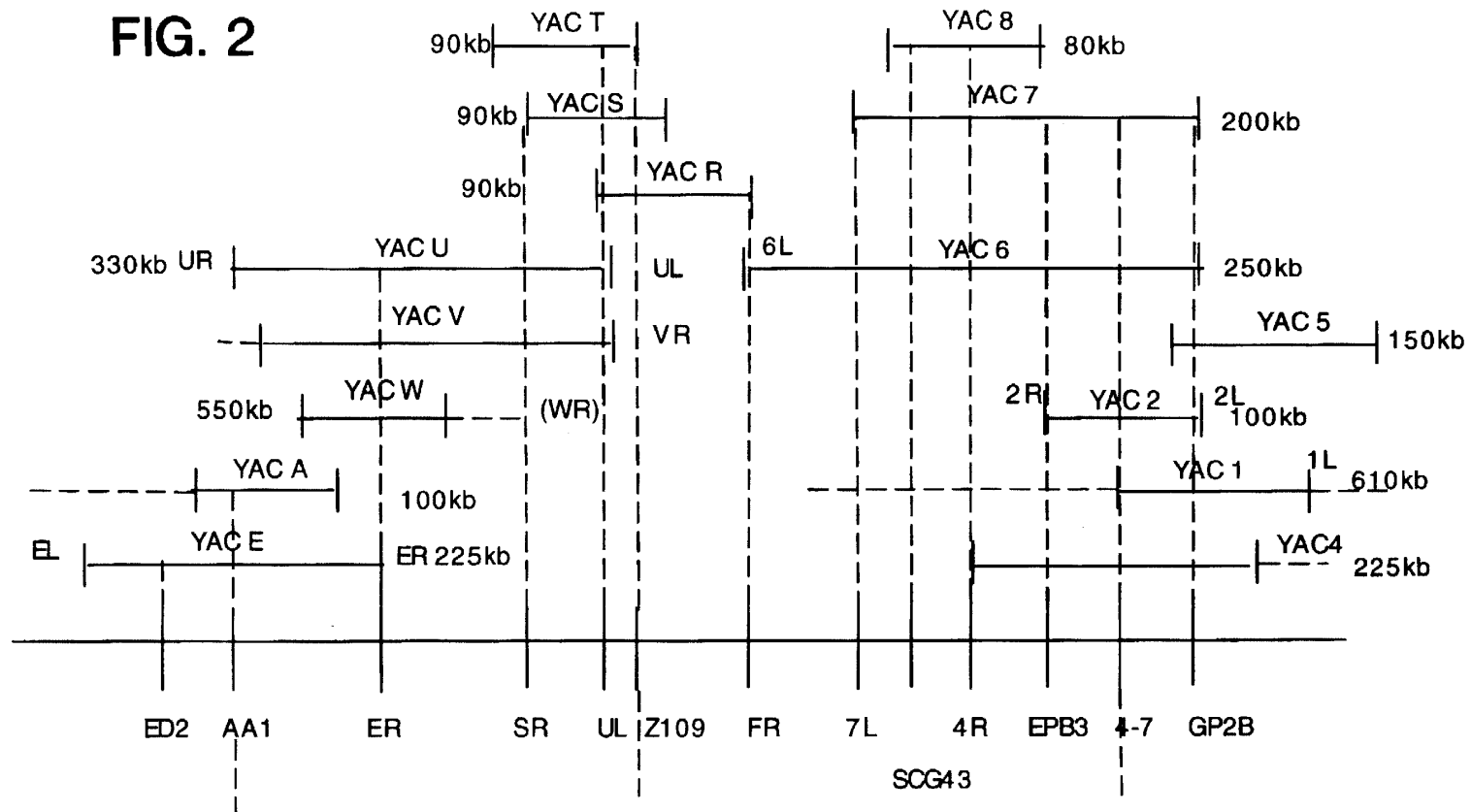
A000147

U.S. Patent

May 5, 1998

Sheet 1 of 18

5,747,282



U.S. Patent

May 5, 1998

Sheet 2 of 18

5,747,282

A000148

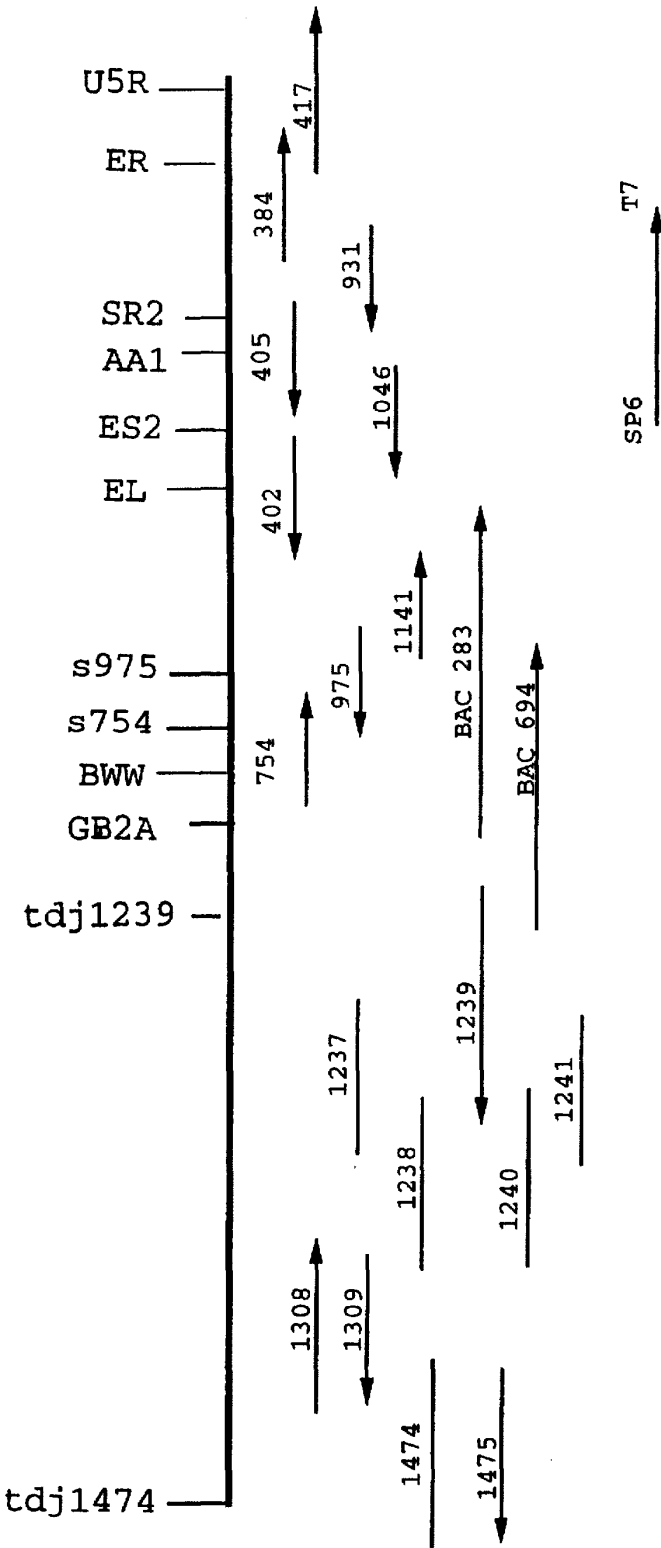
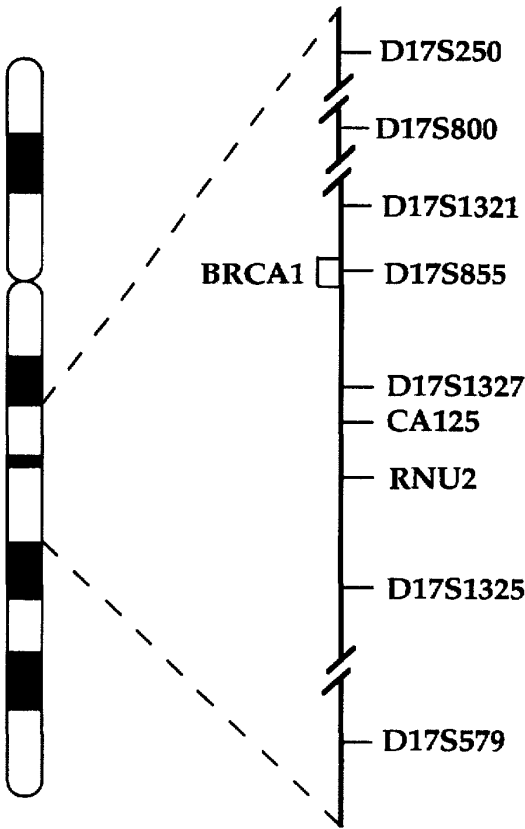


FIG. 3

FIG. 4



SEQ. ID
NO:

82 BRCA1
83 RPT1
84 RIN1
85 RFP1
C3HC4 motif

CPICLELIKEPVSTK-CDHIFCKFCMLKLLNQKK---GPSQCPLCK
CPICLELLKEPVSAD-CNHSFCRACITLNYESNRNTDGKGNCPVCR
CPICLDMLKNTMTTKECLHRFCSDCIVTALRS-----GNKECPTCR
CPVCLQYFAEPMMLD-CGHNICCACLARCWGTAC---TNVSCPQCR
C--C-----C-H--C--C-----C--C

FIG. 5

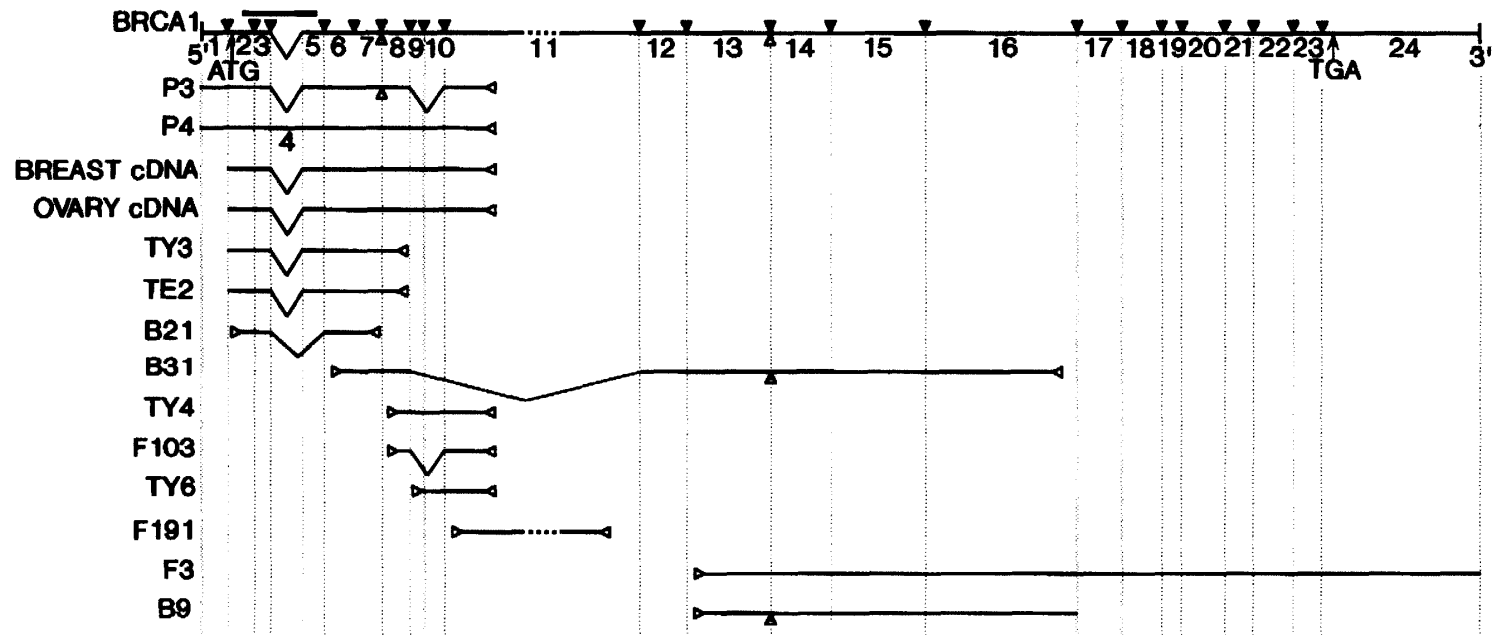


FIG. 6

U.S. Patent

May 5, 1998

Sheet 5 of 18

5,747,282

A000151

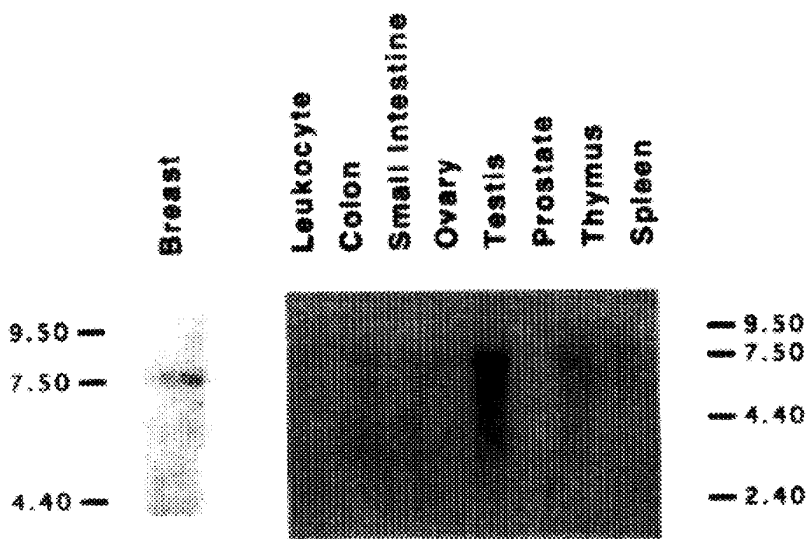


FIG. 7

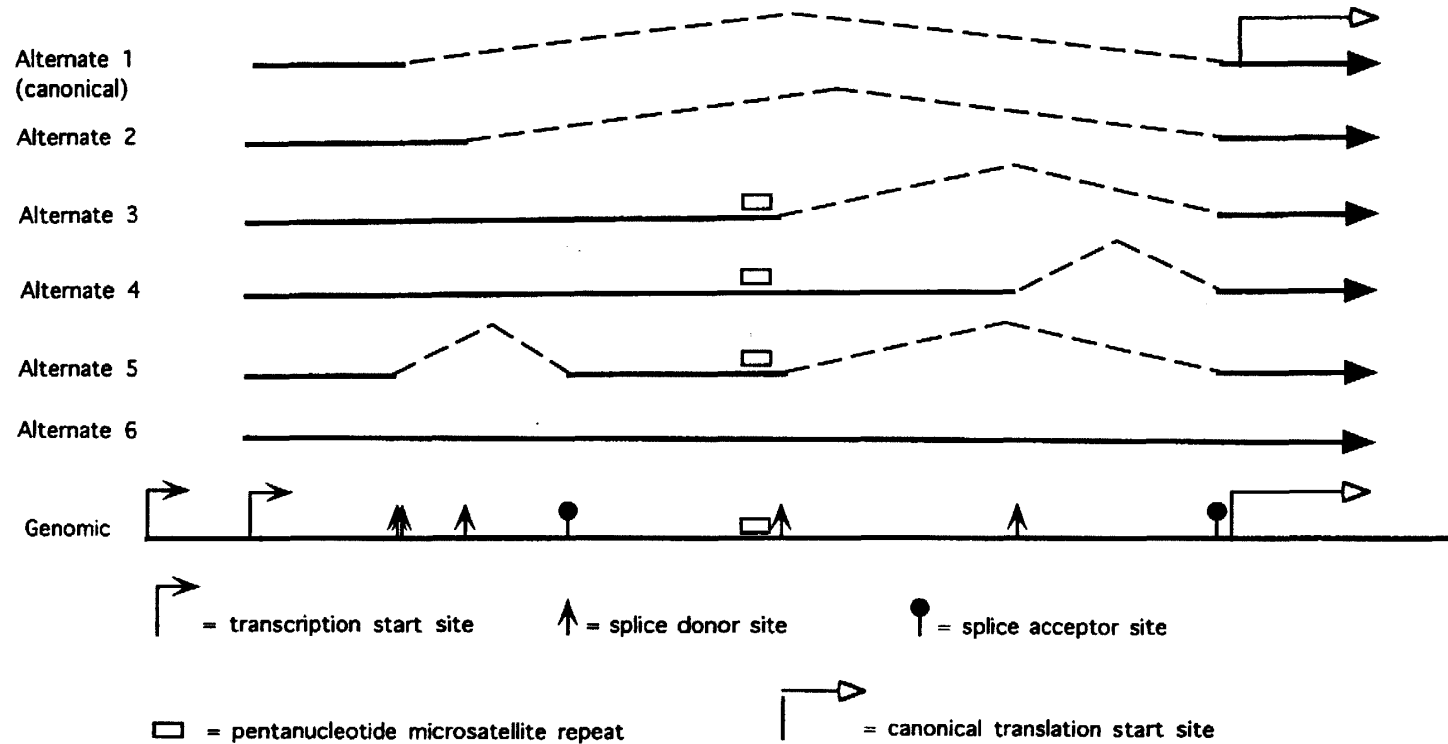


FIG. 8

U.S. Patent

May 5, 1998

Sheet 7 of 18

5,747,282

U.S. Patent

May 5, 1998

Sheet 8 of 18

5,747,282

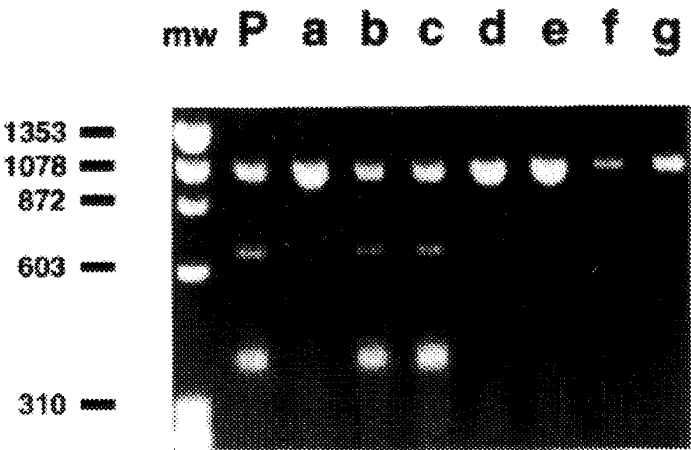


FIG. 9A

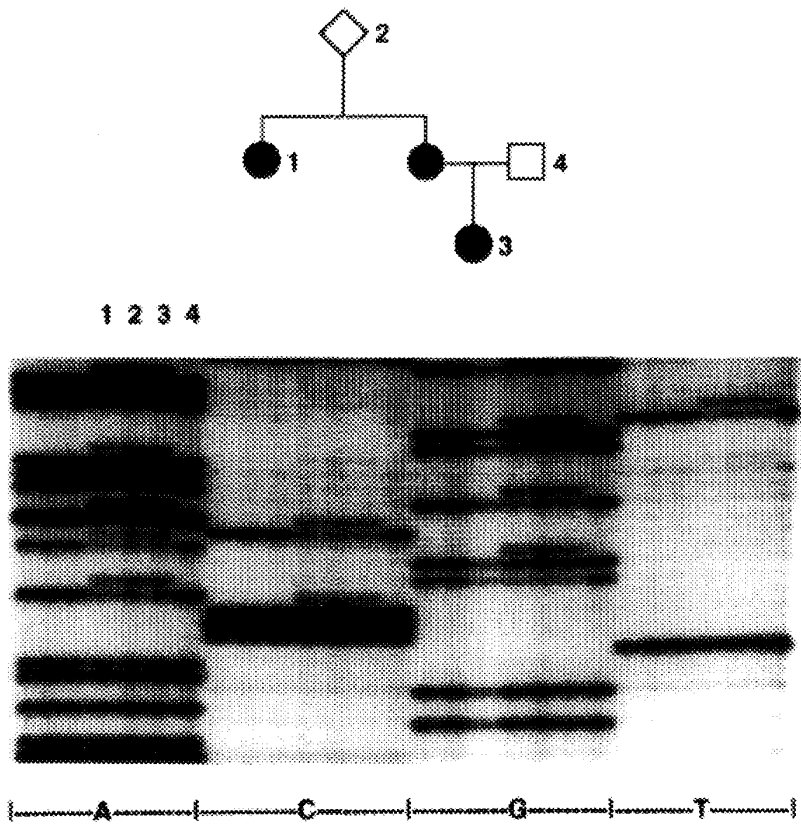


FIG. 9B

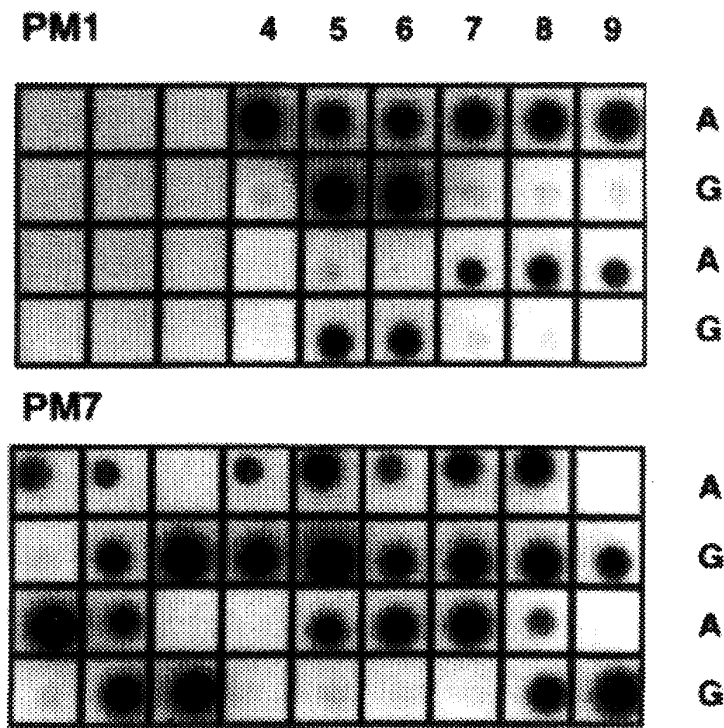
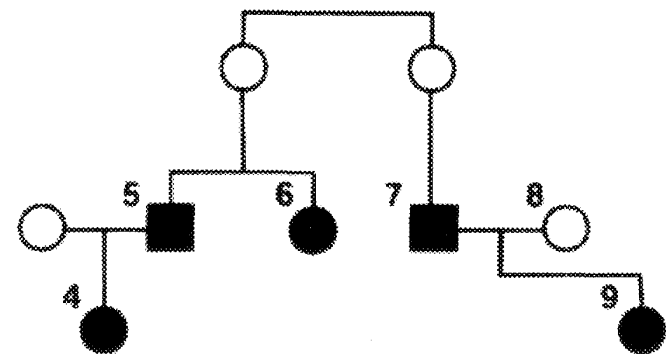


FIG. 9C

A000157

A000158

A000159

A000160

May 5, 1998

Sheet 15 of 18

5,747,282

12301 GGCTTTAATATGTAAAAGTGAAGAGTTCACCTCCAAATCAGTAGAGAGTAATATTGAAGG
12301 CCAAATATTTGGGAAAACCTATCGGAAGAAGGCAAGCCTCCCCAAGCTTAAGCCATGTAAAC
12361 TGAAAATCTAATTATAGGAGCATTGTACTGAGCCACAGATAATACAAGAGCGTCCCCT
12421 CACAAATAAATTAAAGCGTAAAAGGAGACCTACATCAGGCCTTCATCCTGAGGATTTTAT
12481 CAAGAAAGCAGATTTGGCAGTTCAAAAGACTCCTGAAATGATAAATCAGGGAACTAACCA
12541 AACGGAGCAGAATGGTCAAGTGATGAATATTACTAATAGTGGTCATGAGAATAAAACAAA
12601 AGGTGATTTCTATTGAGAATGAGAAAAATCCTAACCCAATAGAATCACTCGAAAAAGAATC
12661 TGCTTTCAAAACGAAAGCTGAACCTATAAGCAGCAGTATAAGCAATATGGAACCTCGAATT
12721 AAATATCCACAATTCAAAGCACCTAAAAAGAATAGGCTGAGGAGGAAGTCTTCTACCAG
12781 GCATATTCATGCGCTTGAACCTAGTAGTCAGTAGAAATCTAAGCCCACCTAATTGTACTGA
12841 ATTGCAAATTGATAGTTGTTCTAGCAGTGAAGAGATAAGAAAAAAAAGTACCAACCAAT
12901 GCCAGTCAGGCACAGCAGAAACCTACAACCTATGGAAAGGTAAAGAAACCTGCCTACTGGAGC
12961 CAAGAAGAGTAACAAGCCAAATGAACAGACAAGTAAAAGACATGACAGCGTACTTTCCC
13021 AGAGCTGAAGTTAACAAATGCACCTGGTTCTTTTACTAAGTGTTCAAATACCAGTGAACCT
13081 TAAAGAATTTGTCAATCCTAGCCTTCCAAGAGAAGAAAAAGAAGAGAACTAGAAAACAGT
13141 TAAAGTGCTAATAATGTGTAAGACCCCCAAGATCTCATGTTAAGTGGAGAAAGGGTTTT
13201 GCAAACCTGAAAGATCTGTAGAGAGTAGCAGTATTTCAATGGTACCTGGTACTGATTATGG
13261 CAATCAGGAAAGTATCTCGTTACTGGAAGTTAGCACTCTAGGGAAGGCCAAAAACAGAACC
13321 AAATAAATGTGTGAGTCAGTGTGCAGCATTGTGAAAACCCCAAGGGACTAATTCATGGTTG
13381 TTCCAAAGATAATAGAAATGACACAGAAGGCTTTAAGTATCCATTGGGACATGAAGTTAA
13441 CCACAGTCGGGGAACAAGCATAGAAATGGAAGAAAGTGAACCTTGATGCTCAGTATTTGCA
13501 GAATACATTCAAGGTTTCAAAGCGCCAGTCATTGTGCTCAGTTTCAAATCCAGGAAATGC
13561 AGAAGAGGAATGTGCAACATTCTCTGCCACTCTGGGTCTTAAAGAAACAAAGTCCAAA
13621 AGTCATCTTTGAAATGTGAACAAAAGGAAGAAATCAAGGAAAGAATGAGTCTAATATCAA
13681 GCCTGTACAGCAGTTAATATCTACTGCAGGCTTTCTGTGGTTGGTCAGAAAGATAAGCC
13741 AGTTGATAATGCCAAATGTAGTATCAAAGAGGCTTCTAGGTTTTGTCTCATCTCAGTT
13801 CAGAGGCAACGAAACTGGACTCATTACTCCAAATAAAACATGGACTTTTACAAAAACCCATA
13861 TCGTATACCACCACCTTTTCCCATCAAGTCATTGTGTTAAAACTAAATGTAAGAAAAATCT
13921 GCTAGAGGAAAACCTTTGAGGAACATTCAATGTCACCTGAAAGAGAAATGGGAAATGAGAA
13981 CATCCAAGTACAGTGAGCACAATTAGCCGTAATAACATTAGAGAAAATGTTTTTAAAGA
14041 AGCCAGCTCAAGCAATATTAATGAAGTAGGTTCCAGTACTAATGAAGTGGCTCCAGTAT
14101 TAATGAAATAGGTTCCAGTGATGAAAACATTCAAGCAGAACTAGGTAGAAACAGAGGGCC
14161 AAAATTGAATGCTATGCTTAGATTAGGGGTTTTTGCAACCTGAGGCTTATAAAACAAAGTCT
14221 TCCTGGAAGTAATTGTAAGCATCCTGAAATAAAAAAGCAAGAATATGAAGAAGTAGTTCA
14281 GACTGTTAATACAGATTTCTCTCCATATCTGATTTAGATAACTTAGAACAGCCTATGGG
14341 AAGTAGTCATGCATCTCAGGTTTGTCTGAGACACCTGATGACCTGTTAGATGATGGTGA
14401 AATAAAGGAAGATACTAGTATTTGTCTGAAATGACATTAAAGGAAAGTTCTGCTGTTTTAG
14461 CAAAAGCGTCCAGAAAGGAGAGCTTAGCAGGAGTCCTAGCCCTTTACCCATACACATTT
14521 GGCTCAGGGTTACCGAAGAGGGGCCAAGAAATAGAGTCTCAGAAGAGAACTTATCTAG
14581 TGAGGATGAAGAGCTTCCCTGCTTCCAACACTTGTTATTTGGTAAAGTAAACAATATACC
14641 TTCTCAGTCTACTAGGCATAGCACCGTTGCTACCGAGTGTCTGTCTAAGAACACAGAGGA
14701 GAATTTATTATCATTGAAGAATAGCTTAAATGACTGCAGTAACCAGGTAATATTGGCAA
14761 GGCATCTCAGGAACATCACCTTAGTGAGGAAACAAAATGTTCTGCTAGCTTGTTTTCTTC
14821 ACAGTGCAGTGAATTGGAAGACTTGACTGCAAAATACAAACACCCAGGATCCTTTCTTGAT
14881 TGGTTCTTCCAAACAAATGAGGCATCAGTCTGAAAGCCAGGGAGTTGGTCTGAGTGACAA
14941 GGAATTGGTTTCAGATGATGAAGAAAGAGGAACGGGCTTGGAAGAAAAATAATCAAGAAGA
15001 GCAAAGCATGGATTCAAACCTTAGGtatttgaaccagggtttttgtgtttgccccagtcctat
15061 ttatagaagtgaagctaaatgttttatgcttttggggagcacatttacaattttccaagta
15121 tagttaaaggaaagctgttctttaaacttgaaacattgttctctcctaagggtgtcttctcataga
15181 aaaaagtccttcacacagctaggagcgtctcttgactgaatgagcttttaacatccta
15241 tactggtggaacttactctgtgtttcattttataaaqcaaatccgggtgtcccaaaqcaat

FIG. 10E

A000162

A000163

A000164

5,747,282

1

**17Q-LINKED BREAST AND OVARIAN
CANCER SUSCEPTIBILITY GENE**

**CROSS REFERENCE TO RELATED
APPLICATIONS**

This application is a continuation-in-part of application Ser. No. 08/409,305 filed on 24 Mar. 1995, abandoned, which is a continuation-in-part of application Ser. No. 08/348,824 filed on 29 Nov. 1994, abandoned, which is a continuation-in-part of application Ser. No. 08/308,104 filed on 16 Sep. 1994, which is a continuation-in-part of application Ser. No. 08/300,266, filed on 2 Sep. 1994, abandoned, which is a continuation-in-part of application Ser. No. 08/289,221, filed on 12 Aug. 1994, abandoned, all incorporated herein by reference.

FIELD OF THE INVENTION

The present invention relates generally to the field of human genetics. Specifically, the present invention relates to methods and materials used to isolate and detect a human breast and ovarian cancer predisposing gene (BRCA1), some mutant alleles of which cause susceptibility to cancer, in particular, breast and ovarian cancer. More specifically, the invention relates to germline mutations in the BRCA1 gene and their use in the diagnosis of predisposition to breast and ovarian cancer. The present invention further relates to somatic mutations in the BRCA1 gene in human breast and ovarian cancer and their use in the diagnosis and prognosis of human breast and ovarian cancer. Additionally, the invention relates to somatic mutations in the BRCA1 gene in other human cancers and their use in the diagnosis and prognosis of human cancers. The invention also relates to the therapy of human cancers which have a mutation in the BRCA1 gene, including gene therapy, protein replacement therapy and protein mimetics. The invention further relates to the screening of drugs for cancer therapy. Finally, the invention relates to the screening of the BRCA1 gene for mutations, which are useful for diagnosing the predisposition to breast and ovarian cancer.

The publications and other materials used herein to illuminate the background of the invention, and in particular, cases to provide additional details respecting the practice, are incorporated herein by reference, and for convenience, are referenced by author and date in the following text and respectively grouped in the appended List of References.

BACKGROUND OF THE INVENTION

The genetics of cancer is complicated, involving multiple dominant, positive regulators of the transformed state (oncogenes) as well as multiple recessive, negative regulators (tumor suppressor genes). Over one hundred oncogenes have been characterized. Fewer than a dozen tumor suppressor genes have been identified, but the number is expected to increase beyond fifty (Knudson, 1993).

The involvement of so many genes underscores the complexity of the growth control mechanisms that operate in cells to maintain the integrity of normal tissue. This complexity is manifest in another way. So far, no single gene has been shown to participate in the development of all, or even the majority of human cancers. The most common oncogenic mutations are in the H-ras gene, found in 10–15% of all solid tumors (Anderson et al., 1992). The most frequently mutated tumor suppressor genes are the TP53 gene, homozygously deleted in roughly 50% of all tumors, and CDKN2, which was homozygously deleted in 46% of tumor

2

cell lines examined (Kamb et al., 1994). Without a target that is common to all transformed cells, the dream of a “magic bullet” that can destroy or revert cancer cells while leaving normal tissue unharmed is improbable. The hope for a new generation of specifically targeted antitumor drugs may rest on the ability to identify tumor suppressor genes or oncogenes that play general roles in control of cell division.

The tumor suppressor genes which have been cloned and characterized influence susceptibility to: 1) Retinoblastoma (RB1); 2) Wilms’ tumor (WT1); 3) Li-Fraumeni (TP53); 4) Familial adenomatous polyposis (APC); 5) Neurofibromatosis type 1 (NF1); 6) Neurofibromatosis type 2 (NF2); 7 von Hippel-Lindau syndrome (VHL); 8) Multiple endocrine neoplasia type 2A (MEN2A); and 9) Melanoma (CDKN2).

Tumor suppressor loci that have been mapped genetically but not yet isolated include genes for: Multiple endocrine neoplasia type 1 (MEN1); Lynch cancer family syndrome 2 (LCFS2); Neuroblastoma (NB); Basal cell nevus syndrome (BCNS); Beckwith-Wiedemann syndrome (BWS); Renal cell carcinoma (RCC); Tuberous sclerosis 1 (TSC1); and Tuberous sclerosis 2 (TSC2). The tumor suppressor genes that have been characterized to date encode products with similarities to a variety of protein types, including DNA binding proteins (WT1), ancillary transcription regulators (RB1), GTPase activating proteins or GAPs (NF1), cytoskeletal components (NF2), membrane bound receptor kinases (MEN2A), cell cycle regulators (CDKN2) and others with no obvious similarity to known proteins (APC and VHL).

In many cases, the tumor suppressor gene originally identified through genetic studies has been shown to be lost or mutated in some sporadic tumors. This result suggests that regions of chromosomal aberration may signify the position of important tumor suppressor genes involved both in genetic predisposition to cancer and in sporadic cancer.

One of the hallmarks of several tumor suppressor genes characterized to date is that they are deleted at high frequency in certain tumor types. The deletions often involve loss of a single allele, a so-called loss of heterozygosity (LOH), but may also involve homozygous deletion of both alleles. For LOH, the remaining allele is presumed to be nonfunctional, either because of a preexisting inherited mutation, or because of a secondary sporadic mutation.

Breast cancer is one of the most significant diseases that affects women. At the current rate, American women have a 1 in 8 risk of developing breast cancer by age 95 (American Cancer Society, 1992). Treatment of breast cancer at later stages is often futile and disfiguring, making early detection a high priority in medical management of the disease. Ovarian cancer, although less frequent than breast cancer is often rapidly fatal and is the fourth most common cause of cancer mortality in American women. Genetic factors contribute to an ill-defined proportion of breast cancer incidence, estimated to be about 5% of all cases but approximately 25% of cases diagnosed before age 40 (Claus et al., 1991). Breast cancer has been subdivided into two types, early-age onset and late-age onset, based on an inflection in the age-specific incidence curve around age 50. Mutation of one gene, BRCA1, is thought to account for approximately 45% of familial breast cancer, but at least 80% of families with both breast and ovarian cancer (Easton et al., 1993).

Intense efforts to isolate the BRCA1 gene have proceeded since it was first mapped in 1990 (Hall et al., 1990; Narod et al., 1991). A second locus, BRCA2, has recently been mapped to chromosome 13q (Wooster et al., 1994) and appears to account for a proportion of early-onset breast cancer roughly equal to BRCA1, but confers a lower risk of

ovarian cancer. The remaining susceptibility to early-onset breast cancer is divided between as yet unmapped genes for familial cancer, and rarer germline mutations in genes such as TP53 (Malkin et al., 1990). It has also been suggested that heterozygote carriers for defective forms of the Ataxia-Telangiectasia gene are at higher risk for breast cancer (Swift et al., 1976; Swift et al., 1991). Late-age onset breast cancer is also often familial although the risks in relatives are not as high as those for early-onset breast cancer (Cannon-Albright et al., 1994; Mettlin et al., 1990). However, the percentage of such cases due to genetic susceptibility is unknown.

Breast cancer has long been recognized to be, in part, a familial disease (Anderson, 1972). Numerous investigators have examined the evidence for genetic inheritance and concluded that the data are most consistent with dominant inheritance for a major susceptibility locus or loci (Bishop and Gardner, 1980; Go et al., 1983; Williams and Anderson, 1984; Bishop et al., 1988; Newman et al., 1988; Claus et al., 1991). Recent results demonstrate that at least three loci exist which convey susceptibility to breast cancer as well as other cancers. These loci are the TP53 locus on chromosome 17p (Malkin et al., 1990), a 17q-linked susceptibility locus known as BRCA1 (Hall et al., 1990), and one or more loci responsible for the unmapped residual. Hall et al. (1990) indicated that the inherited breast cancer susceptibility in kindreds with early age onset is linked to chromosome 17q21; although subsequent studies by this group using a more appropriate genetic model partially refuted the limitation to early onset breast cancer (Margaritte et al., 1992).

Most strategies for cloning the 17q-linked breast cancer predisposing gene (BRCA1) require precise genetic localization studies. The simplest model for the functional role of BRCA1 holds that alleles of BRCA1 that predispose to cancer are recessive to wild type alleles; that is, cells that contain at least one wild type BRCA1 allele are not cancerous. However, cells that contain one wild type BRCA1 allele and one predisposing allele may occasionally suffer loss of the wild type allele either by random mutation or by chromosome loss during cell division (nondisjunction). All the progeny of such a mutant cell lack the wild type function of BRCA1 and may develop into tumors. According to this model, predisposing alleles of BRCA1 are recessive, yet susceptibility to cancer is inherited in a dominant fashion: women who possess one predisposing allele (and one wild type allele) risk developing cancer, because their mammary epithelial cells may spontaneously lose the wild type BRCA1 allele. This model applies to a group of cancer susceptibility loci known as tumor suppressors or antioncogenes, a class of genes that includes the retinoblastoma gene and neurofibromatosis gene. By inference this model may also explain the BRCA1 function, as has recently been suggested (Smith et al., 1992).

A second possibility is that BRCA1 predisposing alleles are truly dominant; that is, a wild type allele of BRCA1 cannot overcome the tumor forming role of the predisposing allele. Thus, a cell that carries both wild type and mutant alleles would not necessarily lose the wild type copy of BRCA1 before giving rise to malignant cells. Instead, mammary cells in predisposed individuals would undergo some other stochastic change(s) leading to cancer.

If BRCA1 predisposing alleles are recessive, the BRCA1 gene is expected to be expressed in normal mammary tissue but not functionally expressed in mammary tumors. In contrast, if BRCA1 predisposing alleles are dominant, the wild type BRCA1 gene may or may not be expressed in normal mammary tissue. However, the predisposing allele will likely be expressed in breast tumor cells.

The 17q linkage of BRCA1 was independently confirmed in three of five kindreds with both breast cancer and ovarian cancer (Narod et al., 1991). These studies claimed to localize the gene within a very large region, 15 centiMorgans (cM), or approximately 15 million base pairs, to either side of the linked marker pCMM86 (D17S74). However, attempts to define the region further by genetic studies, using markers surrounding pCMM86, proved unsuccessful. Subsequent studies indicated that the gene was considerably more proximal (Easton et al., 1993) and that the original analysis was flawed (Margaritte et al., 1992). Hall et al., (1992) recently localized the BRCA1 gene to an approximately 8 cM interval (approximately 8 million base pairs) bounded by Mfd15 (D17S250) on the proximal side and the human GIP gene on the distal side. A slightly narrower interval for the BRCA1 locus, based on publicly available data, was agreed upon at the Chromosome 17 workshop in March of 1992 (Fain, 1992). The size of these regions and the uncertainty associated with them has made it exceedingly difficult to design and implement physical mapping and/or cloning strategies for isolating the BRCA1 gene.

Identification of a breast cancer susceptibility locus would permit the early detection of susceptible individuals and greatly increase our ability to understand the initial steps which lead to cancer. As susceptibility loci are often altered during tumor progression, cloning these genes could also be important in the development of better diagnostic and prognostic products, as well as better cancer therapies.

SUMMARY OF THE INVENTION

The present invention relates generally to the field of human genetics. Specifically, the present invention relates to methods and materials used to isolate and detect a human breast cancer predisposing gene (BRCA1), some alleles of which cause susceptibility to cancer, in particular breast and ovarian cancer. More specifically, the present invention relates to germline mutations in the BRCA1 gene and their use in the diagnosis of predisposition to breast and ovarian cancer. The invention further relates to somatic mutations in the BRCA1 gene in human breast cancer and their use in the diagnosis and prognosis of human breast and ovarian cancer. Additionally, the invention relates to somatic mutations in the BRCA1 gene in other human cancers and their use in the diagnosis and prognosis of human cancers. The invention also relates to the therapy of human cancers which have a mutation in the BRCA1 gene, including gene therapy, protein replacement therapy and protein mimetics. The invention further relates to the screening of drugs for cancer therapy. Finally, the invention relates to the screening of the BRCA1 gene for mutations, which are useful for diagnosing the predisposition to breast and ovarian cancer.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 is a diagram showing the order of loci neighboring BRCA1 as determined by the chromosome 17 workshop. FIG. 1 is reproduced from Fain, 1992.

FIG. 2 is a schematic map of YACs which define part of Mfd15-Mfd188 region.

FIG. 3 is a schematic map of STSs, PIs and BACs in the BRCA1 region.

FIG. 4 is a schematic map of human chromosome 17. The pertinent region containing BRCA1 is expanded to indicate the relative positions of two previously identified genes, CA125 and RNU2. BRCA1 spans the marker D17S855.

FIG. 5 shows alignment of the BRCA1 zinc-finger domain with 3 other zinc-finger domains that scored highest

in a Smith-Waterman alignment. RPT1 encodes a protein that appears to be a negative regulator of the IL-2 receptor in mouse. RIN1 encodes a DNA-binding protein that includes a RING-finger motif related to the zinc-finger. RFP1 encodes a putative transcription factor that is the N-terminal domain of the RET oncogene product. The bottom line contains the C3HC4 consensus zinc-finger sequence showing the positions of cysteines and one histidine that form the zinc ion binding pocket.

FIG. 6 is a diagram of BRCA1 mRNA showing the locations of introns and the variants of BRCA1 mRNA produced by alternative splicing. Intron locations are shown by dark triangles and the exons are numbered below the line representing the cDNA. The top cDNA is the composite used to generate the peptide sequence of BRCA1. Alternative forms identified as cDNA clones or hybrid selection clones are shown below.

FIG. 7 shows the tissue expression pattern of BRCA1. The blot was obtained from Clontech and contains RNA from the indicated tissues. Hybridization conditions were as recommended by the manufacturer using a probe consisting of nucleotide positions 3631 to 3930 of BRCA1. Note that both breast and ovary are heterogeneous tissues and the percentage of relevant epithelial cells can be variable. Molecular weight standards are in kilobases.

FIG. 8 is a diagram of the 5' untranslated region plus the beginning of the translated region of BRCA1 showing the locations of introns and the variants of BRCA1 mRNA produced by alternative splicing. Intron locations are shown by broken dashed lines. Six alternate splice forms are shown.

FIG. 9A shows a nonsense mutation in Kindred 2082. P indicates the person originally screened, b and c are haplotype carriers, a, d, e, f, and g do not carry the BRCA1 haplotype. The C to T mutation results in a stop codon and creates a site for the restriction enzyme AvrII. PCR amplification products are cut with this enzyme. The carriers are heterozygous for the site and therefore show three bands. Non-carriers remain uncut.

FIG. 9B shows a mutation and cosegregation analysis in BRCA1 kindreds. Carrier individuals are represented as filled circles and squares in the pedigree diagrams. Frameshift mutation in Kindred 1910. The first three lanes are control, noncarrier samples. Lanes labeled 1-3 contain sequences from carrier individuals. Lane 4 contains DNA from a kindred member who does not carry the BRCA1 mutation. The diamond is used to prevent identification of the kindred. The frameshift resulting from the additional C is apparent in lanes labeled 1, 2, and 3.

FIG. 9C shows a mutation and cosegregation analysis in BRCA1 kindreds. Carrier individuals are represented as filled circles and squares in the pedigree diagrams. Inferred regulatory mutation in Kindred 2035. ASO analysis of carriers and noncarriers of 2 different polymorphisms (PM1 and PM7) which were examined for heterozygosity in the germline and compared to the heterozygosity of lymphocyte mRNA. The top 2 rows of each panel contain PCR products amplified from genomic DNA and the bottom 2 rows contain PCR products amplified from cDNA. "A" and "G" are the two alleles detected by the ASO. The dark spots indicate that a particular allele is present in the sample. The first three lanes of PM7 represent the three genotypes in the general population.

FIG. 10A-10H show genomic sequence of BRCA1. The lower case letters denote intron sequence while the upper case letters denote exon sequence. Indefinite intervals within

introns are designated with vvvvvvvvvvvv. Known polymorphic sites are shown as underlined and boldface type.

DETAILED DESCRIPTION OF THE INVENTION

The present invention relates generally to the field of human genetics. Specifically, the present invention relates to methods and materials used to isolate and detect a human breast cancer predisposing gene (BRCA1), some alleles of which cause susceptibility to cancer, in particular breast and ovarian cancer. More specifically, the present invention relates to germline mutations in the BRCA1 gene and their use in the diagnosis of predisposition to breast and ovarian cancer. The invention further relates to somatic mutations in the BRCA1 gene in human breast cancer and their use in the diagnosis and prognosis of human breast and ovarian cancer. Additionally, the invention relates to somatic mutations in the BRCA1 gene in other human cancers and their use in the diagnosis and prognosis of human cancers. The invention also relates to the therapy of human cancers which have a mutation in the BRCA1 gene, including gene therapy, protein replacement therapy and protein mimetics. The invention further relates to the screening of drugs for cancer therapy. Finally, the invention relates to the screening of the BRCA1 gene for mutations, which are useful for diagnosing the predisposition to breast and ovarian cancer.

The present invention provides an isolated polynucleotide comprising all, or a portion of the BRCA1 locus or of a mutated BRCA1 locus, preferably at least eight bases and not more than about 100 kb in length. Such polynucleotides may be antisense polynucleotides. The present invention also provides a recombinant construct comprising such an isolated polynucleotide, for example, a recombinant construct suitable for expression in a transformed host cell.

Also provided by the present invention are methods of detecting a polynucleotide comprising a portion of the BRCA1 locus or its expression product in an analyte. Such methods may further comprise the step of amplifying the portion of the BRCA1 locus, and may further include a step of providing a set of polynucleotides which are primers for amplification of said portion of the BRCA1 locus. The method is useful for either diagnosis of the predisposition to cancer or the diagnosis or prognosis of cancer.

The present invention also provides isolated antibodies, preferably monoclonal antibodies, which specifically bind to an isolated polypeptide comprised of at least five amino acid residues encoded by the BRCA1 locus.

The present invention also provides kits for detecting in an analyte a polynucleotide comprising a portion of the BRCA1 locus, the kits comprising a polynucleotide complementary to the portion of the BRCA1 locus packaged in a suitable container, and instructions for its use.

The present invention further provides methods of preparing a polynucleotide comprising polymerizing nucleotides to yield a sequence comprised of at least eight consecutive nucleotides of the BRCA1 locus; and methods of preparing a polypeptide comprising polymerizing amino acids to yield a sequence comprising at least five amino acids encoded within the BRCA1 locus.

The present invention further provides methods of screening the BRCA1 gene to identify mutations. Such methods may further comprise the step of amplifying a portion of the BRCA1 locus, and may further include a step of providing a set of polynucleotides which are primers for amplification of said portion of the BRCA1 locus. The method is useful for identifying mutations for use in either diagnosis of the predisposition to cancer or the diagnosis or prognosis of cancer.

5,747,282

7

The present invention further provides methods of screening suspected BRCA1 mutant alleles to identify mutations in the BRCA1 gene.

In addition, the present invention provides methods of screening drugs for cancer therapy to identify suitable drugs for restoring BRCA1 gene product function.

Finally, the present invention provides the means necessary for production of gene-based therapies directed at cancer cells. These therapeutic agents may take the form of polynucleotides comprising all or a portion of the BRCA1 locus placed in appropriate vectors or delivered to target cells in more direct ways such that the function of the BRCA1 protein is reconstituted. Therapeutic agents may also take the form of polypeptides based on either a portion of, or the entire protein sequence of BRCA1. These may functionally replace the activity of BRCA1 in vivo.

It is a discovery of the present invention that the BRCA1 locus which predisposes individuals to breast cancer and ovarian cancer, is a gene encoding a BRCA1 protein, which has been found to have no significant homology with known protein or DNA sequences. This gene is termed BRCA1 herein. It is a discovery of the present invention that mutations in the BRCA1 locus in the germline are indicative of a predisposition to breast cancer and ovarian cancer. Finally, it is a discovery of the present invention that somatic mutations in the BRCA1 locus are also associated with breast cancer, ovarian cancer and other cancers, which represents an indicator of these cancers or of the prognosis of these cancers. The mutational events of the BRCA1 locus can involve deletions, insertions and point mutations within the coding sequence and the non-coding sequence.

Starting from a region on the long arm of human chromosome 17 of the human genome, 17q, which has a size estimated at about 8 million base pairs, a region which contains a genetic locus, BRCA1, which causes susceptibility to cancer, including breast and ovarian cancer, has been identified.

The region containing the BRCA1 locus was identified using a variety of genetic techniques. Genetic mapping techniques initially defined the BRCA1 region in terms of recombination with genetic markers. Based upon studies of large extended families ("kindreds") with multiple cases of breast cancer (and ovarian cancer cases in some kindreds), a chromosomal region has been pinpointed that contains the BRCA1 gene as well as other putative susceptibility alleles in the BRCA1 locus. Two meiotic breakpoints have been discovered on the distal side of the BRCA1 locus which are expressed as recombinants between genetic markers and the disease, and one recombinant on the proximal side of the BRCA1 locus. Thus, a region which contains the BRCA1 locus is physically bounded by these markers.

The use of the genetic markers provided by this invention allowed the identification of clones which cover the region from a human yeast artificial chromosome (YAC) or a human bacterial artificial chromosome (BAC) library. It also allowed for the identification and preparation of more easily manipulated cosmid, P1 and BAC clones from this region and the construction of a contig from a subset of the clones. These cosmids, P1s, YACs and BACs provide the basis for cloning the BRCA1 locus and provide the basis for developing reagents effective, for example, in the diagnosis and treatment of breast and/or ovarian cancer. The BRCA1 gene and other potential susceptibility genes have been isolated from this region. The isolation was done using software trapping (a computational method for identifying sequences likely to contain coding exons, from contiguous or discon-

8

tinuous genomic DNA sequences), hybrid selection techniques and direct screening, with whole or partial cDNA inserts from cosmids, P1s and BACs, in the region to screen cDNA libraries. These methods were used to obtain sequences of loci expressed in breast and other tissue. These candidate loci were analyzed to identify sequences which confer cancer susceptibility. We have discovered that there are mutations in the coding sequence of the BRCA1 locus in kindreds which are responsible for the 17q-linked cancer susceptibility known as BRCA1. This gene was not known to be in this region. The present invention not only facilitates the early detection of certain cancers, so vital to patient survival, but also permits the detection of susceptible individuals before they develop cancer.

Population Resources

Large, well-documented Utah kindreds are especially important in providing good resources for human genetic studies. Each large kindred independently provides the power to detect whether a BRCA1 susceptibility allele is segregating in that family. Recombinants informative for localization and isolation of the BRCA1 locus could be obtained only from kindreds large enough to confirm the presence of a susceptibility allele. Large sibships are especially important for studying breast cancer, since penetrance of the BRCA1 susceptibility allele is reduced both by age and sex, making informative sibships difficult to find. Furthermore, large sibships are essential for constructing haplotypes of deceased individuals by inference from the haplotypes of their close relatives.

While other populations may also provide beneficial information, such studies generally require much greater effort, and the families are usually much smaller and thus less informative. Utah's age-adjusted breast cancer incidence is 20% lower than the average U.S. rate. The lower incidence in Utah is probably due largely to an early age at first pregnancy, increasing the probability that cases found in Utah kindreds carry a genetic predisposition.

Genetic Mapping

Given a set of informative families, genetic markers are essential for linking a disease to a region of a chromosome. Such markers include restriction fragment length polymorphisms (RFLPs) (Botstein et al., 1980), markers with a variable number of tandem repeats (VNTRs) (Jeffreys et al., 1985; Nakamura et al., 1987), and an abundant class of DNA polymorphisms based on short tandem repeats (STRs), especially repeats of CpA (Weber and May, 1989; Litt et al., 1989). To generate a genetic map, one selects potential genetic markers and tests them using DNA extracted from members of the kindreds being studied.

Genetic markers useful in searching for a genetic locus associated with a disease can be selected on an ad hoc basis, by densely covering a specific chromosome, or by detailed analysis of a specific region of a chromosome. A preferred method for selecting genetic markers linked with a disease involves evaluating the degree of informativeness of kindreds to determine the ideal distance between genetic markers of a given degree of polymorphism, then selecting markers from known genetic maps which are ideally spaced for maximal efficiency. Informativeness of kindreds is measured by the probability that the markers will be heterozygous in unrelated individuals. It is also most efficient to use STR markers which are detected by amplification of the target nucleic acid sequence using PCR; such markers are highly informative, easy to assay (Weber and May, 1989), and can be assayed simultaneously using multiplexing strategies (Skolnick and Wallace, 1988), greatly reducing the number of experiments required.

Once linkage has been established, one needs to find markers that flank the disease locus, i.e., one or more markers proximal to the disease locus, and one or more markers distal to the disease locus. Where possible, candidate markers can be selected from a known genetic map. Where none is known, new markers can be identified by the STR technique, as shown in the Examples.

Genetic mapping is usually an iterative process. In the present invention, it began by defining flanking genetic markers around the BRCA1 locus, then replacing these flanking markers with other markers that were successively closer to the BRCA1 locus. As an initial step, recombination events, defined by large extended kindreds, helped specifically to localize the BRCA1 locus as either distal or proximal to a specific genetic marker (Goldgar et al., 1994).

The region surrounding BRCA1, until the disclosure of the present invention, was not well mapped and there were few markers. Therefore, short repetitive sequences on cosmids subcloned from YACs, which had been physically mapped, were analyzed in order to develop new genetic markers. Using this approach, one marker of the present invention, 42D6, was discovered which replaced pCMM86 as the distal flanking marker for the BRCA1 region. Since 42D6 is approximately 14 cM from pCMM86, the BRCA1 region was thus reduced by approximately 14 centiMorgans (Easton et al., 1993). The present invention thus began by finding a much more closely linked distal flanking marker of the BRCA1 region. BRCA1 was then discovered to be distal to the genetic marker Mfd15. Therefore, BRCA1 was shown to be in a region of 6 to 10 million bases bounded by Mfd15 and 42D6. Marker Mfd191 was subsequently discovered to be distal to Mfd15 and proximal to BRCA1. Thus, Mfd15 was replaced with Mfd191 as the closest proximal genetic marker. Similarly, it was discovered that genetic marker Mfd188 could replace genetic marker 42D6, narrowing the region containing the BRCA1 locus to approximately 1.5 million bases. Then the marker Mfd191 was replaced with tdj1474 as the proximal marker and Mfd188 was replaced with USR as the distal marker, further narrowing the BRCA1 region to a small enough region to allow isolation and characterization of the BRCA1 locus (see FIG. 3), using techniques known in the art and described herein.

Physical Mapping

Three distinct methods were employed to physically map the region. The first was the use of yeast artificial chromosomes (YACs) to clone the region which is flanked by tdj1474 and USR. The second was the creation of a set of P1, BAC and cosmid clones which cover the region containing the BRCA1 locus.

Yeast Artificial Chromosomes (YACs). Once a sufficiently small region containing the BRCA1 locus was identified, physical isolation of the DNA in the region proceeded by identifying a set of overlapping YACs which covers the region. Useful YACs can be isolated from known libraries, such as the St. Louis and CEPH YAC libraries, which are widely distributed and contain approximately 50,000 YACs each. The YACs isolated were from these publicly accessible libraries and can be obtained from a number of sources including the Michigan Genome Center. Clearly, others who had access to these YACs, without the disclosure of the present invention, would not have known the value of the specific YACs we selected since they would not have known which YACs were within, and which YACs outside of, the smallest region containing the BRCA1 locus.

Cosmid, P1 and BAC Clones. In the present invention, it is advantageous to proceed by obtaining cosmid, P1, and BAC clones to cover this region. The smaller size of these

inserts, compared to YAC inserts, makes them more useful as specific hybridization probes. Furthermore, having the cloned DNA in bacterial cells, rather than in yeast cells, greatly increases the ease with which the DNA of interest can be manipulated, and improves the signal-to-noise ratio of hybridization assays. For cosmid subclones of YACs, the DNA is partially digested with the restriction enzyme Sau3A and cloned into the BamHI site of the pWE15 cosmid vector (Stratagene, cat. #1251201). The cosmids containing human sequences are screened by hybridization with human repetitive DNA (e.g., Gibco/BRL, Human C₀t-1 DNA, cat. 5279SA), and then fingerprinted by a variety of techniques, as detailed in the Examples.

P1 and BAC clones are obtained by screening libraries constructed from the total human genome with specific sequence tagged sites (STSs) derived from the YACs, cosmids or P1s and BACs, isolated as described herein.

These P1, BAC and cosmid clones can be compared by interspersed repetitive sequence (IRS) PCR and/or restriction enzyme digests followed by gel electrophoresis and comparison of the resulting DNA fragments ("fingerprints") (Maniatis et al., 1982). The clones can also be characterized by the presence of STSs. The fingerprints are used to define an overlapping contiguous set of clones which covers the region but is not excessively redundant, referred to herein as a "minimum tiling path". Such a minimum tiling path forms the basis for subsequent experiments to identify cDNAs which may originate from the BRCA1 locus.

Coverage of the Gap with P1 and BAC Clones. To cover any gaps in the BRCA1 contig between the identified cosmids with genomic clones, clones in P1 and BAC vectors which contain inserts of genomic DNA roughly twice as large as cosmids for P1s and still greater for BACs (Sternberg, 1990; Sternberg et al., 1990; Pierce et al., 1992; Shizuya et al., 1992) were used. P1 clones were isolated by Genome Sciences using PCR primers provided by us for screening. BACs were provided by hybridization techniques in Dr. Mel Simon's laboratory. The strategy of using P1 clones also permitted the covering of the genomic region with an independent set of clones not derived from YACs. This guards against the possibility of other deletions in YACs that have not been detected. These new sequences derived from the P1 clones provide the material for further screening for candidate genes, as described below.

Gene Isolation.

There are many techniques for testing genomic clones for the presence of sequences likely to be candidates for the coding sequence of a locus one is attempting to isolate, including but not limited to:

- a. zoo blots
- b. identifying HTF islands
- c. exon trapping
- d. hybridizing cDNA to cosmids or YACs.
- e. screening cDNA libraries.

(a) Zoo blots. The first technique is to hybridize cosmids to Southern blots to identify DNA sequences which are evolutionarily conserved, and which therefore give positive hybridization signals with DNA from species of varying degrees of relationship to humans (such as monkey, cow, chicken, pig, mouse and rat). Southern blots containing such DNA from a variety of species are commercially available (Clontech, Cat. 7753-1).

(b) Identifying HTF islands. The second technique involves finding regions rich in the nucleotides C and G, which often occur near or within coding sequences. Such sequences are called HTF (HpaI tiny fragment) or CpG islands, as restric-

tion enzymes specific for sites which contain CpG dimers cut frequently in these regions (Lindsay et al., 1987).

(c) Exon trapping. The third technique is exon trapping, a method that identifies sequences in genomic DNA which contain splice junctions and therefore are likely to comprise coding sequences of genes. Exon amplification (Buckler et al., 1991) is used to select and amplify exons from DNA clones described above. Exon amplification is based on the selection of RNA sequences which are flanked by functional 5' and/or 3' splice sites. The products of the exon amplification are used to screen the breast cDNA libraries to identify a manageable number of candidate genes for further study. Exon trapping can also be performed on small segments of sequenced DNA using computer programs or by software trapping.

(d) Hybridizing cDNA to Cosmids, PIs, BACs or YACs. The fourth technique is a modification of the selective enrichment technique which utilizes hybridization of cDNA to cosmids, PIs, BACs or YACs and permits transcribed sequences to be identified in, and recovered from cloned genomic DNA (Kandpal et al. 1990). The selective enrichment technique, as modified for the present purpose, involves binding DNA from the region of BRCA1 present in a YAC to a column matrix and selecting cDNAs from the relevant libraries which hybridize with the bound DNA, followed by amplification and purification of the bound DNA, resulting in a great enrichment for cDNAs in the region represented by the cloned genomic DNA.

(e) Identification of cDNAs. The fifth technique is to identify cDNAs that correspond to the BRCA1 locus. Hybridization probes containing putative coding sequences, selected using any of the above techniques, are used to screen various libraries, including breast tissue cDNA libraries, ovarian cDNA libraries, and any other necessary libraries.

Another variation on the theme of direct selection of cDNA was also used to find candidate genes for BRCA1 (Lovett et al., 1991; Futreal, 1993). This method uses cosmid, P1 or BAC DNA as the probe. The probe DNA is digested with a blunt cutting restriction enzyme such as HaeIII. Double stranded adapters are then ligated onto the DNA and serve as binding sites for primers in subsequent PCR amplification reactions using biotinylated primers. Target cDNA is generated from mRNA derived from tissue samples, e.g., breast tissue, by synthesis of either random primed or oligo(dT) primed first strand followed by second strand synthesis. The cDNA ends are rendered blunt and ligated onto double-stranded adapters. These adapters serve as amplification sites for PCR. The target and probe sequences are denatured and mixed with human C₀t-1 DNA to block repetitive sequences.

Solution hybridization is carried out to high C₀t-1/2 values to ensure hybridization of rare target cDNA molecules. The annealed material is then captured on avidin beads, washed at high stringency and the retained cDNAs are eluted and amplified by PCR. The selected cDNA is subjected to further rounds of enrichment before cloning into a plasmid vector for analysis.

Testing the cDNA for Candidacy

Proof that the cDNA is the BRCA1 locus is obtained by finding sequences in DNA extracted from affected kindred members which create abnormal BRCA1 gene products or abnormal levels of BRCA1 gene product. Such BRCA1 susceptibility alleles will co-segregate with the disease in large kindreds. They will also be present at a much higher frequency in non-kindred individuals with breast and ovarian cancer than in individuals in the general population.

Finally, since tumors often mutate somatically at loci which are in other instances mutated in the germline, we expect to see normal germline BRCA1 alleles mutated into sequences which are identical or similar to BRCA1 susceptibility alleles in DNA extracted from tumor tissue. Whether one is comparing BRCA1 sequences from tumor tissue to BRCA1 alleles from the germline of the same individuals, or one is comparing germline BRCA1 alleles from cancer cases to those from unaffected individuals, the key is to find mutations which are serious enough to cause obvious disruption to the normal function of the gene product. These mutations can take a number of forms. The most severe forms would be frame shift mutations or large deletions which would cause the gene to code for an abnormal protein or one which would significantly alter protein expression. Less severe disruptive mutations would include small in-frame deletions and nonconservative base pair substitutions which would have a significant effect on the protein produced, such as changes to or from a cysteine residue, from a basic to an acidic amino acid or vice versa, from a hydrophobic to hydrophilic amino acid or vice versa, or other mutations which would affect secondary, tertiary or quaternary protein structure. Silent mutations or those resulting in conservative amino acid substitutions would not generally be expected to disrupt protein function.

According to the diagnostic and prognostic method of the present invention, alteration of the wild-type BRCA1 locus is detected. In addition, the method can be performed by detecting the wild-type BRCA1 locus and confirming the lack of a predisposition to cancer at the BRCA1 locus. "Alteration of a wild-type gene" encompasses all forms of mutations including deletions, insertions and point mutations in the coding and noncoding regions. Deletions may be of the entire gene or of only a portion of the gene. Point mutations may result in stop codons, frameshift mutations or amino acid substitutions. Somatic mutations are those which occur only in certain tissues, e.g., in the tumor tissue, and are not inherited in the germline. Germline mutations can be found in any of a body's tissues and are inherited. If only a single allele is somatically mutated, an early neoplastic state is indicated. However, if both alleles are somatically mutated, then a late neoplastic state is indicated. The finding of BRCA1 mutations thus provides both diagnostic and prognostic information. A BRCA1 allele which is not deleted (e.g., found on the sister chromosome to a chromosome carrying a BRCA1 deletion) can be screened for other mutations, such as insertions, small deletions, and point mutations. It is believed that many mutations found in tumor tissues will be those leading to decreased expression of the BRCA1 gene product. However, mutations leading to non-functional gene products would also lead to a cancerous state. Point mutational events may occur in regulatory regions, such as in the promoter of the gene, leading to loss or diminution of expression of the mRNA. Point mutations may also abolish proper RNA processing, leading to loss of expression of the BRCA1 gene product, or to a decrease in mRNA stability or translation efficiency.

Useful diagnostic techniques include, but are not limited to fluorescent in situ hybridization (FISH), direct DNA sequencing, PFGE analysis, Southern blot analysis, single stranded conformation analysis (SSCA), RNase protection assay, allele-specific oligonucleotide (ASO), dot blot analysis and PCR-SSCP, as discussed in detail further below.

Predisposition to cancers, such as breast and ovarian cancer, and the other cancers identified herein, can be ascertained by testing any tissue of a human for mutations of the BRCA1 gene. For example, a person who has inherited

5,747,282

13

a germline BRCA1 mutation would be prone to develop cancers. This can be determined by testing DNA from any tissue of the person's body. Most simply, blood can be drawn and DNA extracted from the cells of the blood. In addition, prenatal diagnosis can be accomplished by testing fetal cells, placental cells or amniotic cells for mutations of the BRCA1 gene. Alteration of a wild-type BRCA1 allele, whether, for example, by point mutation or deletion, can be detected by any of the means discussed herein.

There are several methods that can be used to detect DNA sequence variation. Direct DNA sequencing, either manual sequencing or automated fluorescent sequencing can detect sequence variation. For a gene as large as BRCA1, manual sequencing is very labor-intensive, but under optimal conditions, mutations in the coding sequence of a gene are rarely missed. Another approach is the single-stranded conformation polymorphism assay (SSCA) (Orita et al., 1989). This method does not detect all sequence changes, especially if the DNA fragment size is greater than 200 bp, but can be optimized to detect most DNA sequence variation. The reduced detection sensitivity is a disadvantage, but the increased throughput possible with SSCA makes it an attractive, viable alternative to direct sequencing for mutation detection on a research basis. The fragments which have shifted mobility on SSCA gels are then sequenced to determine the exact nature of the DNA sequence variation. Other approaches based on the detection of mismatches between the two complementary DNA strands include clamped denaturing gel electrophoresis (CDGE) (Sheffield et al., 1991), heteroduplex analysis (HA) (White et al., 1992) and chemical mismatch cleavage (CMC) (Grompe et al., 1989). None of the methods described above will detect large deletions, duplications or insertions, nor will they detect a regulatory mutation which affects transcription or translation of the protein. Other methods which might detect these classes of mutations such as a protein truncation assay or the asymmetric assay, detect only specific types of mutations and would not detect missense mutations. A review of currently available methods of detecting DNA sequence variation can be found in a recent review by Grompe (1993). Once a mutation is known, an allele specific detection approach such as allele specific oligonucleotide (ASO) hybridization can be utilized to rapidly screen large numbers of other samples for that same mutation.

In order to detect the alteration of the wild-type BRCA1 gene in a tissue, it is helpful to isolate the tissue free from surrounding normal tissues. Means for enriching tissue preparation for tumor cells are known in the art. For example, the tissue may be isolated from paraffin or cryostat sections. Cancer cells may also be separated from normal cells by flow cytometry. These techniques, as well as other techniques for separating tumor cells from normal cells, are well known in the art. If the tumor tissue is highly contaminated with normal cells, detection of mutations is more difficult.

A rapid preliminary analysis to detect polymorphisms in DNA sequences can be performed by looking at a series of Southern blots of DNA cut with one or more restriction enzymes, preferably with a large number of restriction enzymes. Each blot contains a series of normal individuals and a series of cancer cases, tumors, or both. Southern blots displaying hybridizing fragments (differing in length from control DNA when probed with sequences near or including the BRCA1 locus) indicate a possible mutation. If restriction enzymes which produce very large restriction fragments are used, then pulsed field gel electrophoresis (PFGE) is employed.

14

Detection of point mutations may be accomplished by molecular cloning of the BRCA1 allele(s) and sequencing the allele(s) using techniques well known in the art. Alternatively, the gene sequences can be amplified directly from a genomic DNA preparation from the tumor tissue, using known techniques. The DNA sequence of the amplified sequences can then be determined.

There are six well known methods for a more complete, yet still indirect, test for confirming the presence of a susceptibility allele: 1) single stranded conformation analysis (SSCA) (Orita et al., 1989); 2) denaturing gradient gel electrophoresis (DGGE) (Wartell et al., 1990; Sheffield et al., 1989); 3) RNase protection assays (Finkelstein et al., 1990; Kinszler et al., 1991); 4) allele-specific oligonucleotides (ASOs) (Conner et al., 1983); 5) the use of proteins which recognize nucleotide mismatches, such as the *E. coli* mutS protein (Modrich, 1991); and 6) allele-specific PCR (Rano & Kidd, 1989). For allele-specific PCR, primers are used which hybridize at their 3' ends to a particular BRCA1 mutation. If the particular BRCA1 mutation is not present, an amplification product is not observed. Amplification Refractory Mutation System (ARMS) can also be used, as disclosed in European Patent Application Publication No. 0332435 and in Newton et al., 1989. Insertions and deletions of genes can also be detected by cloning, sequencing and amplification. In addition, restriction fragment length polymorphism (RFLP) probes for the gene or surrounding marker genes can be used to score alteration of an allele or an insertion in a polymorphic fragment. Such a method is particularly useful for screening relatives of an affected individual for the presence of the BRCA1 mutation found in that individual. Other techniques for detecting insertions and deletions as known in the art can be used.

In the first three methods (SSCA, DGGE and RNase protection assay), a new electrophoretic band appears. SSCA detects a band which migrates differentially because the sequence change causes a difference in single-strand, intramolecular base pairing. RNase protection involves cleavage of the mutant polynucleotide into two or more smaller fragments. DGGE detects differences in migration rates of mutant sequences compared to wild-type sequences, using a denaturing gradient gel. In an allele-specific oligonucleotide assay, an oligonucleotide is designed which detects a specific sequence, and the assay is performed by detecting the presence or absence of a hybridization signal. In the mutS assay, the protein binds only to sequences that contain a nucleotide mismatch in a heteroduplex between mutant and wild-type sequences.

Mismatches, according to the present invention, are hybridized nucleic acid duplexes in which the two strands are not 100% complementary. Lack of total homology may be due to deletions, insertions, inversions or substitutions. Mismatch detection can be used to detect point mutations in the gene or in its mRNA product. While these techniques are less sensitive than sequencing, they are simpler to perform on a large number of tumor samples. An example of a mismatch cleavage technique is the RNase protection method. In the practice of the present invention, the method involves the use of a labeled riboprobe which is complementary to the human wild-type BRCA1 gene coding sequence. The riboprobe and either mRNA or DNA isolated from the tumor tissue are annealed (hybridized) together and subsequently digested with the enzyme RNase A which is able to detect some mismatches in a duplex RNA structure. If a mismatch is detected by RNase A, it cleaves at the site of the mismatch. Thus, when the annealed RNA preparation is separated on an electrophoretic gel matrix, if a mismatch

5,747,282

15

has been detected and cleaved by RNase A, an RNA product will be seen which is smaller than the full length duplex RNA for the riboprobe and the mRNA or DNA. The riboprobe need not be the full length of the BRCA1 mRNA or gene but can be a segment of either. If the riboprobe comprises only a segment of the BRCA1 mRNA or gene, it will be desirable to use a number of these probes to screen the whole mRNA sequence for mismatches.

In similar fashion, DNA probes can be used to detect mismatches, through enzymatic or chemical cleavage. See, e.g., Cotton et al., 1988; Shenk et al., 1975; Novack et al., 1986. Alternatively, mismatches can be detected by shifts in the electrophoretic mobility of mismatched duplexes relative to matched duplexes. See, e.g., Cariello, 1988. With either riboprobes or DNA probes, the cellular mRNA or DNA which might contain a mutation can be amplified using PCR (see below) before hybridization. Changes in DNA of the BRCA1 gene can also be detected using Southern hybridization, especially if the changes are gross rearrangements, such as deletions and insertions.

DNA sequences of the BRCA1 gene which have been amplified by use of PCR may also be screened using allele-specific probes. These probes are nucleic acid oligomers, each of which contains a region of the BRCA1 gene sequence harboring a known mutation. For example, one oligomer may be about 30 nucleotides in length, corresponding to a portion of the BRCA1 gene sequence. By use of a battery of such allele-specific probes, PCR amplification products can be screened to identify the presence of a previously identified mutation in the BRCA1 gene. Hybridization of allele-specific probes with amplified BRCA1 sequences can be performed, for example, on a nylon filter. Hybridization to a particular probe under stringent hybridization conditions indicates the presence of the same mutation in the tumor tissue as in the allele-specific probe.

The most definitive test for mutations in a candidate locus is to directly compare genomic BRCA1 sequences from cancer patients with those from a control population. Alternatively, one could sequence messenger RNA after amplification, e.g., by PCR, thereby eliminating the necessity of determining the exon structure of the candidate gene.

Mutations from cancer patients falling outside the coding region of BRCA1 can be detected by examining the non-coding regions, such as introns and regulatory sequences near or within the BRCA1 gene. An early indication that mutations in noncoding regions are important may come from Northern blot experiments that reveal messenger RNA molecules of abnormal size or abundance in cancer patients as compared to control individuals.

Alteration of BRCA1 mRNA expression can be detected by any techniques known in the art. These include Northern blot analysis, PCR amplification and RNase protection. Diminished mRNA expression indicates an alteration of the wild-type BRCA1 gene. Alteration of wild-type BRCA1 genes can also be detected by screening for alteration of wild-type BRCA1 protein. For example, monoclonal antibodies immunoreactive with BRCA1 can be used to screen a tissue. Lack of cognate antigen would indicate a BRCA1 mutation. Antibodies specific for products of mutant alleles could also be used to detect mutant BRCA1 gene product. Such immunological assays can be done in any convenient formats known in the art. These include Western blots, immunohistochemical assays and ELISA assays. Any means for detecting an altered BRCA1 protein can be used to detect alteration of wild-type BRCA1 genes. Functional assays, such as protein binding determinations, can be used. In

16

addition, assays can be used which detect BRCA1 biochemical function. Finding a mutant BRCA1 gene product indicates alteration of a wild-type BRCA1 gene.

Mutant BRCA1 genes or gene products can also be detected in other human body samples, such as serum, stool, urine and sputum. The same techniques discussed above for detection of mutant BRCA1 genes or gene products in tissues can be applied to other body samples. Cancer cells are sloughed off from tumors and appear in such body samples. In addition, the BRCA1 gene product itself may be secreted into the extracellular space and found in these body samples even in the absence of cancer cells. By screening such body samples, a simple early diagnosis can be achieved for many types of cancers. In addition, the progress of chemotherapy or radiotherapy can be monitored more easily by testing such body samples for mutant BRCA1 genes or gene products.

The methods of diagnosis of the present invention are applicable to any tumor in which BRCA1 has a role in tumorigenesis. The diagnostic method of the present invention is useful for clinicians, so they can decide upon an appropriate course of treatment.

The primer pairs of the present invention are useful for determination of the nucleotide sequence of a particular BRCA1 allele using PCR. The pairs of single-stranded DNA primers can be annealed to sequences within or surrounding the BRCA1 gene on chromosome 17q21 in order to prime amplifying DNA synthesis of the BRCA1 gene itself. A complete set of these primers allows synthesis of all of the nucleotides of the BRCA1 gene coding sequences, i.e., the exons. The set of primers preferably allows synthesis of both intron and exon sequences. Allele-specific primers can also be used. Such primers anneal only to particular BRCA1 mutant alleles, and thus will only amplify a product in the presence of the mutant allele as a template.

In order to facilitate subsequent cloning of amplified sequences, primers may have restriction enzyme site sequences appended to their 5' ends. Thus, all nucleotides of the primers are derived from BRCA1 sequences or sequences adjacent to BRCA1, except for the few nucleotides necessary to form a restriction enzyme site. Such enzymes and sites are well known in the art. The primers themselves can be synthesized using techniques which are well known in the art. Generally, the primers can be made using oligonucleotide synthesizing machines which are commercially available. Given the sequence of the BRCA1 open reading frame shown in SEQ ID NO: 1, design of particular primers is well within the skill of the art.

The nucleic acid probes provided by the present invention are useful for a number of purposes. They can be used in Southern hybridization to genomic DNA and in the RNase protection method for detecting point mutations already discussed above. The probes can be used to detect PCR amplification products. They may also be used to detect mismatches with the BRCA1 gene or mRNA using other techniques.

It has been discovered that individuals with the wild-type BRCA1 gene do not have cancer which results from the BRCA1 allele. However, mutations which interfere with the function of the BRCA1 protein are involved in the pathogenesis of cancer. Thus, the presence of an altered (or a mutant) BRCA1 gene which produces a protein having a loss of function, or altered function, directly correlates to an increased risk of cancer. In order to detect a BRCA1 gene mutation, a biological sample is prepared and analyzed for a difference between the sequence of the BRCA1 allele being analyzed and the sequence of the wild-type BRCA1

A000172

5,747,282

17

allele. Mutant BRCA1 alleles can be initially identified by any of the techniques described above. The mutant alleles are then sequenced to identify the specific mutation of the particular mutant allele. Alternatively, mutant BRCA1 alleles can be initially identified by identifying mutant (altered) BRCA1 proteins, using conventional techniques. The mutant alleles are then sequenced to identify the specific mutation for each allele. The mutations, especially those which lead to an altered function of the BRCA1 protein, are then used for the diagnostic and prognostic methods of the present invention.

DEFINITIONS

The present invention employs the following definitions:

"Amplification of Polynucleotides" utilizes methods such as the polymerase chain reaction (PCR), ligation amplification (or ligase chain reaction, LCR) and amplification methods based on the use of Q-beta replicase. These methods are well known and widely practiced in the art. See, e.g., U.S. Pat. Nos. 4,683,195 and 4,683,202 and Innis et al., 1990 (for PCR); and Wu et al., 1989a (for LCR). Reagents and hardware for conducting PCR are commercially available. Primers useful to amplify sequences from the BRCA1 region are preferably complementary to, and hybridize specifically to sequences in the BRCA1 region or in regions that flank a target region therein. BRCA1 sequences generated by amplification may be sequenced directly. Alternatively, but less desirably, the amplified sequence(s) may be cloned prior to sequence analysis. A method for the direct cloning and sequence analysis of enzymatically amplified genomic segments has been described by Scharf, 1986.

"Analyte polynucleotide" and "analyte strand" refer to a single- or double-stranded polynucleotide which is suspected of containing a target sequence, and which may be present in a variety of types of samples, including biological samples.

"Antibodies." The present invention also provides polyclonal and/or monoclonal antibodies and fragments thereof, and immunologic binding equivalents thereof, which are capable of specifically binding to the BRCA1 polypeptides and fragments thereof or to polynucleotide sequences from the BRCA1 region, particularly from the BRCA1 locus or a portion thereof. The term "antibody" is used both to refer to a homogeneous molecular entity, or a mixture such as a serum product made up of a plurality of different molecular entities. Polypeptides may be prepared synthetically in a peptide synthesizer and coupled to a carrier molecule (e.g., keyhole limpet hemocyanin) and injected over several months into rabbits. Rabbit sera is tested for immunoreactivity to the BRCA1 polypeptide or fragment. Monoclonal antibodies may be made by injecting mice with the protein polypeptides, fusion proteins or fragments thereof. Monoclonal antibodies will be screened by ELISA and tested for specific immunoreactivity with BRCA1 polypeptide or fragments thereof. See, Harlow & Lane, 1988. These antibodies will be useful in assays as well as pharmaceuticals.

Once a sufficient quantity of desired polypeptide has been obtained, it may be used for various purposes. A typical use is the production of antibodies specific for binding. These antibodies may be either polyclonal or monoclonal, and may be produced by in vitro or in vivo techniques well known in the art. For production of polyclonal antibodies, an appropriate target immune system, typically mouse or rabbit, is selected. Substantially purified antigen is presented to the immune system in a fashion determined by methods appropriate for the animal and by other parameters well known to

18

immunologists. Typical sites for injection are in footpads, intramuscularly, intraperitoneally, or intradermally. Of course, other species may be substituted for mouse or rabbit. Polyclonal antibodies are then purified using techniques known in the art, adjusted for the desired specificity.

An immunological response is usually assayed with an immunoassay. Normally, such immunoassays involve some purification of a source of antigen, for example, that produced by the same cells and in the same fashion as the antigen. A variety of immunoassay methods are well known in the art. See, e.g., Harlow & Lane, 1988, or Goding, 1986.

Monoclonal antibodies with affinities of 10^{-8} M⁻¹ or preferably 10^{-9} to 10^{-10} M⁻¹ or stronger will typically be made by standard procedures as described, e.g., in Harlow & Lane, 1988 or Goding, 1986. Briefly, appropriate animals will be selected and the desired immunization protocol followed. After the appropriate period of time, the spleens of such animals are excised and individual spleen cells fused, typically, to immortalized myeloma cells under appropriate selection conditions. Thereafter, the cells are clonally separated and the supernatants of each clone tested for their production of an appropriate antibody specific for the desired region of the antigen.

Other suitable techniques involve in vitro exposure of lymphocytes to the antigenic polypeptides, or alternatively, to selection of libraries of antibodies in phage or similar vectors. See Huse et al., 1989. The polypeptides and antibodies of the present invention may be used with or without modification. Frequently, polypeptides and antibodies will be labeled by joining, either covalently or non-covalently, a substance which provides for a detectable signal. A wide variety of labels and conjugation techniques are known and are reported extensively in both the scientific and patent literature. Suitable labels include radionuclides, enzymes, substrates, cofactors, inhibitors, fluorescent agents, chemiluminescent agents, magnetic particles and the like. Patents teaching the use of such labels include U.S. Pat. Nos. 3,817,837; 3,850,752; 3,939,350; 3,996,345; 4,277,437; 4,275,149 and 4,366,241. Also, recombinant immunoglobulins may be produced (see U.S. Pat. No. 4,816,567).

"Binding partner" refers to a molecule capable of binding a ligand molecule with high specificity, as for example, an antigen and an antigen-specific antibody or an enzyme and its inhibitor. In general, the specific binding partners must bind with sufficient affinity to immobilize the analyte copy/complementary strand duplex (in the case of polynucleotide hybridization) under the isolation conditions. Specific binding partners are known in the art and include, for example, biotin and avidin or streptavidin, IgG and protein A, the numerous, known receptor-ligand couples, and complementary polynucleotide strands. In the case of complementary polynucleotide binding partners, the partners are normally at least about 15 bases in length, and may be at least 40 bases in length. The polynucleotides may be composed of DNA, RNA, or synthetic nucleotide analogs.

A "biological sample" refers to a sample of tissue or fluid suspected of containing an analyte polynucleotide or polypeptide from an individual including, but not limited to, e.g., plasma, serum, spinal fluid, lymph fluid, the external sections of the skin, respiratory, intestinal, and genitourinary tracts, tears, saliva, blood cells, tumors, organs, tissue and samples of in vitro cell culture constituents.

As used herein, the terms "diagnosing" or "prognosing," as used in the context of neoplasia, are used to indicate 1) the classification of lesions as neoplasia, 2) the determination of the severity of the neoplasia, or 3) the monitoring of the disease progression, prior to, during and after treatment.

A000173

5,747,282

19

"Encode". A polynucleotide is said to "encode" a polypeptide if, in its native state or when manipulated by methods well known to those skilled in the art, it can be transcribed and/or translated to produce the mRNA for and/or the polypeptide or a fragment thereof. The anti-sense strand is the complement of such a nucleic acid, and the encoding sequence can be deduced therefrom.

"Isolated" or "substantially pure". An "isolated" or "substantially pure" nucleic acid (e.g., an RNA, DNA or a mixed polymer) is one which is substantially separated from other cellular components which naturally accompany a native human sequence or protein. e.g., ribosomes, polymerases, many other human genome sequences and proteins. The term embraces a nucleic acid sequence or protein which has been removed from its naturally occurring environment, and includes recombinant or cloned DNA isolates and chemically synthesized analogs or analogs biologically synthesized by heterologous systems.

"BRCA1 Allele" refers to normal alleles of the BRCA1 locus as well as alleles carrying variations that predispose individuals to develop cancer of many sites including, for example, breast, ovarian, colorectal and prostate cancer. Such predisposing alleles are also called "BRCA1 susceptibility alleles".

"BRCA1 Locus," "BRCA1 Gene," "BRCA1 Nucleic Acids" or "BRCA1 Polynucleotide" each refer to polynucleotides, all of which are in the BRCA1 region, that are likely to be expressed in normal tissue, certain alleles of which predispose an individual to develop breast, ovarian, colorectal and prostate cancers. Mutations at the BRCA1 locus may be involved in the initiation and/or progression of other types of tumors. The locus is indicated in part by mutations that predispose individuals to develop cancer. These mutations fall within the BRCA1 region described infra. The BRCA1 locus is intended to include coding sequences, intervening sequences and regulatory elements controlling transcription and/or translation. The BRCA1 locus is intended to include all allelic variations of the DNA sequence.

These terms, when applied to a nucleic acid, refer to a nucleic acid which encodes a BRCA1 polypeptide, fragment, homolog or variant, including, e.g., protein fusions or deletions. The nucleic acids of the present invention will possess a sequence which is either derived from, or substantially similar to a natural BRCA1-encoding gene or one having substantial homology with a natural BRCA1-encoding gene or a portion thereof. The coding sequence for a BRCA1 polypeptide is shown in SEQ ID NO:1, with the amino acid sequence shown in SEQ ID NO:2.

The polynucleotide compositions of this invention include RNA, cDNA, genomic DNA, synthetic forms, and mixed polymers, both sense and antisense strands, and may be chemically or biochemically modified or may contain non-natural or derivatized nucleotide bases, as will be readily appreciated by those skilled in the art. Such modifications include, for example, labels, methylation, substitution of one or more of the naturally occurring nucleotides with an analog, internucleotide modifications such as uncharged linkages (e.g., methyl phosphonates, phosphotriesters, phosphoamidates, carbamates, etc.), charged linkages (e.g., phosphorothioates, phosphorodithioates, etc.), pendent moieties (e.g., polypeptides), intercalators (e.g., acridine, psoralen, etc.), chelators, alkylators, and modified linkages (e.g., alpha anomeric nucleic acids, etc.). Also included are synthetic molecules that mimic polynucleotides in their ability to bind

20

to a designated sequence via hydrogen bonding and other chemical interactions. Such molecules are known in the art and include, for example, those in which peptide linkages substitute for phosphate linkages in the backbone of the molecule.

The present invention provides recombinant nucleic acids comprising all or part of the BRCA1 region. The recombinant construct may be capable of replicating autonomously in a host cell. Alternatively, the recombinant construct may become integrated into the chromosomal DNA of the host cell. Such a recombinant polynucleotide comprises a polynucleotide of genomic, cDNA, semi-synthetic, or synthetic origin which, by virtue of its origin or manipulation, 1) is not associated with all or a portion of a polynucleotide with which it is associated in nature; 2) is linked to a polynucleotide other than that to which it is linked in nature; or 3) does not occur in nature.

Therefore, recombinant nucleic acids comprising sequences otherwise not naturally occurring are provided by this invention. Although the wild-type sequence may be employed, it will often be altered, e.g., by deletion, substitution or insertion.

cDNA or genomic libraries of various types may be screened as natural sources of the nucleic acids of the present invention, or such nucleic acids may be provided by amplification of sequences resident in genomic DNA or other natural sources, e.g., by PCR. The choice of cDNA libraries normally corresponds to a tissue source which is abundant in mRNA for the desired proteins. Phage libraries are normally preferred, but other types of libraries may be used. Clones of a library are spread onto plates, transferred to a substrate for screening, denatured and probed for the presence of desired sequences.

The DNA sequences used in this invention will usually comprise at least about five codons (15 nucleotides), more usually at least about 7-15 codons, and most preferably, at least about 35 codons. One or more introns may also be present. This number of nucleotides is usually about the minimal length required for a successful probe that would hybridize specifically with a BRCA1-encoding sequence.

Techniques for nucleic acid manipulation are described generally, for example, in Sambrook et al., 1989 or Ausubel et al., 1992. Reagents useful in applying such techniques, such as restriction enzymes and the like, are widely known in the art and commercially available from such vendors as New England BioLabs, Boehringer Mannheim, Amersham, Promega Biotec, U. S. Biochemicals, New England Nuclear, and a number of other sources. The recombinant nucleic acid sequences used to produce fusion proteins of the present invention may be derived from natural or synthetic sequences. Many natural gene sequences are obtainable from various cDNA or from genomic libraries using appropriate probes. See, GenBank, National Institutes of Health.

"BRCA1 Region" refers to a portion of human chromosome 17q21 bounded by the markers tdj1474 and U5R. This region contains the BRCA1 locus, including the BRCA1 gene.

As used herein, the terms "BRCA1 locus," "BRCA1 allele" and "BRCA1 region" all refer to the double-stranded DNA comprising the locus, allele, or region, as well as either of the single-stranded DNAs comprising the locus, allele or region.

As used herein, a "portion" of the BRCA1 locus or region or allele is defined as having a minimal size of at least about eight nucleotides, or preferably about 15 nucleotides, or more preferably at least about 25 nucleotides, and may have a minimal size of at least about 40 nucleotides.

A000174

5,747,282

21

"BRCA1 protein" or "BRCA1 polypeptide" refer to a protein or polypeptide encoded by the BRCA1 locus, variants or fragments thereof. The term "polypeptide" refers to a polymer of amino acids and its equivalent and does not refer to a specific length of the product; thus, peptides, oligopeptides and proteins are included within the definition of a polypeptide. This term also does not refer to, or exclude modifications of the polypeptide, for example, glycosylations, acetylations, phosphorylations, and the like. Included within the definition are, for example, polypeptides containing one or more analogs of an amino acid (including, for example, unnatural amino acids, etc.), polypeptides with substituted linkages as well as other modifications known in the art, both naturally and non-naturally occurring. Ordinarily, such polypeptides will be at least about 50% homologous to the native BRCA1 sequence, preferably in excess of about 90%, and more preferably at least about 95% homologous. Also included are proteins encoded by DNA which hybridize under high or low stringency conditions, to BRCA1-encoding nucleic acids and closely related polypeptides or proteins retrieved by antisera to the BRCA1 protein (s).

The length of polypeptide sequences compared for homology will generally be at least about 16 amino acids, usually at least about 20 residues, more usually at least about 24 residues, typically at least about 28 residues, and preferably more than about 35 residues.

"Operably linked" refers to a juxtaposition wherein the components so described are in a relationship permitting them to function in their intended manner. For instance, a promoter is operably linked to a coding sequence if the promoter affects its transcription or expression.

"Probes". Polynucleotide polymorphisms associated with BRCA1 alleles which predispose to certain cancers or are associated with most cancers are detected by hybridization with a polynucleotide probe which forms a stable hybrid with that of the target sequence, under stringent to moderately stringent hybridization and wash conditions. If it is expected that the probes will be perfectly complementary to the target sequence, stringent conditions will be used. Hybridization stringency may be lessened if some mismatching is expected, for example, if variants are expected with the result that the probe will not be completely complementary. Conditions are chosen which rule out nonspecific/adventitious bindings, that is, which minimize noise. Since such indications identify neutral DNA polymorphisms as well as mutations, these indications need further analysis to demonstrate detection of a BRCA1 susceptibility allele.

Probes for BRCA1 alleles may be derived from the sequences of the BRCA1 region or its cDNAs. The probes may be of any suitable length, which span all or a portion of the BRCA1 region, and which allow specific hybridization to the BRCA1 region. If the target sequence contains a sequence identical to that of the probe, the probes may be short, e.g., in the range of about 8–30 base pairs, since the hybrid will be relatively stable under even stringent conditions. If some degree of mismatch is expected with the probe, i.e., if it is suspected that the probe will hybridize to a variant region, a longer probe may be employed which hybridizes to the target sequence with the requisite specificity.

The probes will include an isolated polynucleotide attached to a label or reporter molecule and may be used to isolate other polynucleotide sequences, having sequence similarity by standard methods. For techniques for preparing and labeling probes see, e.g., Sambrook et al., 1989 or

22

Ausubel et al., 1992. Other similar polynucleotides may be selected by using homologous polynucleotides. Alternatively, polynucleotides encoding these or similar polypeptides may be synthesized or selected by use of the redundancy in the genetic code. Various codon substitutions may be introduced, e.g., by silent changes (thereby producing various restriction sites) or to optimize expression for a particular system. Mutations may be introduced to modify the properties of the polypeptide, perhaps to change ligand-binding affinities, interchain affinities, or the polypeptide degradation or turnover rate.

Probes comprising synthetic oligonucleotides or other polynucleotides of the present invention may be derived from naturally occurring or recombinant single- or double-stranded polynucleotides, or be chemically synthesized. Probes may also be labeled by nick translation. Klenow fill-in reaction, or other methods known in the art.

Portions of the polynucleotide sequence having at least about eight nucleotides, usually at least about 15 nucleotides, and fewer than about 6 kb, usually fewer than about 1.0 kb, from a polynucleotide sequence encoding BRCA1 are preferred as probes. The probes may also be used to determine whether mRNA encoding BRCA1 is present in a cell or tissue.

"Protein modifications or fragments" are provided by the present invention for BRCA1 poly-peptides or fragments thereof which are substantially homologous to primary structural sequence but which include, e.g., in vivo or in vitro chemical and biochemical modifications or which incorporate unusual amino acids. Such modifications include, for example, acetylation, carboxylation, phosphorylation, glycosylation, ubiquitination, labeling, e.g., with radionuclides, and various enzymatic modifications, as will be readily appreciated by those well skilled in the art. A variety of methods for labeling polypeptides and of substituents or labels useful for such purposes are well known in the art, and include radioactive isotopes such as ³²P, ligands which bind to labeled antigens (e.g., antibodies), fluorophores, chemiluminescent agents, enzymes, and antigens which can serve as specific binding pair members for a labeled ligand. The choice of label depends on the sensitivity required, ease of conjugation with the primer, stability requirements, and available instrumentation. Methods of labeling polypeptides are well known in the art. See, e.g., Sambrook et al., 1989 or Ausubel et al., 1992.

Besides substantially full-length polypeptides, the present invention provides for biologically active fragments of the polypeptides. Significant biological activities include ligand-binding, immunological activity and other biological activities characteristic of BRCA1 polypeptides. Immunological activities include both immunogenic function in a target immune system, as well as sharing of immunological epitopes for binding, serving as either a competitor or substitute antigen for an epitope of the BRCA1 protein. As used herein, "epitope" refers to an antigenic determinant of a polypeptide. An epitope could comprise three amino acids in a spatial conformation which is unique to the epitope. Generally, an epitope consists of at least five such amino acids, and more usually consists of at least 8–10 such amino acids. Methods of determining the spatial conformation of such amino acids are known in the art.

For immunological purposes, tandem-repeat polypeptide segments may be used as immunogens, thereby producing highly antigenic proteins. Alternatively, such polypeptides will serve as highly efficient competitors for specific bind-

5,747,282

23

ing. Production of antibodies specific for BRCA1 polypeptides or fragments thereof is described below.

The present invention also provides for fusion polypeptides, comprising BRCA1 polypeptides and fragments. Homologous polypeptides may be fusions between two or more BRCA1 polypeptide sequences or between the sequences of BRCA1 and a related protein. Likewise, heterologous fusions may be constructed which would exhibit a combination of properties or activities of the derivative proteins. For example, ligand-binding or other domains may be "swapped" between different new fusion polypeptides or fragments. Such homologous or heterologous fusion polypeptides may display, for example, altered strength or specificity of binding. Fusion partners include immunoglobulins, bacterial β -galactosidase, trpE, protein A, β -lactamase, alpha amylase, alcohol dehydrogenase and yeast alpha mating factor. See, e.g., Godowski et al., 1988.

Fusion proteins will typically be made by either recombinant nucleic acid methods, as described below, or may be chemically synthesized. Techniques for the synthesis of polypeptides are described, for example, in Merrifield, 1963.

"Protein purification" refers to various methods for the isolation of the BRCA1 polypeptides from other biological material, such as from cells transformed with recombinant nucleic acids encoding BRCA1, and are well known in the art. For example, such polypeptides may be purified by immunoaffinity chromatography employing, e.g., the antibodies provided by the present invention. Various methods of protein purification are well known in the art, and include those described in Deutscher, 1990 and Scopes, 1982.

The terms "isolated", "substantially pure", and "substantially homogeneous" are used interchangeably to describe a protein or polypeptide which has been separated from components which accompany it in its natural state. A monomeric protein is substantially pure when at least about 60 to 75% of a sample exhibits a single polypeptide sequence. A substantially pure protein will typically comprise about 60 to 90% W/W of a protein sample, more usually about 95%, and preferably will be over about 99% pure. Protein purity or homogeneity may be indicated by a number of means well known in the art, such as polyacrylamide gel electrophoresis of a protein sample, followed by visualizing a single polypeptide band upon staining the gel. For certain purposes, higher resolution may be provided by using HPLC or other means well known in the art which are utilized for purification.

A BRCA1 protein is substantially free of naturally associated components when it is separated from the native contaminants which accompany it in its natural state. Thus, a polypeptide which is chemically synthesized or synthesized in a cellular system different from the cell from which it naturally originates will be substantially free from its naturally associated components. A protein may also be rendered substantially free of naturally associated components by isolation, using protein purification techniques well known in the art.

A polypeptide produced as an expression product of an isolated and manipulated genetic sequence is an "isolated polypeptide," as used herein, even if expressed in a homologous cell type. Synthetically made forms or molecules expressed by heterologous cells are inherently isolated molecules.

"Recombinant nucleic acid" is a nucleic acid which is not naturally occurring, or which is made by the artificial combination of two otherwise separated segments of sequence. This artificial combination is often accomplished

24

by either chemical synthesis means, or by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques. Such is usually done to replace a codon with a redundant codon encoding the same or a conservative amino acid, while typically introducing or removing a sequence recognition site. Alternatively, it is performed to join together nucleic acid segments of desired functions to generate a desired combination of functions.

"Regulatory sequences" refers to those sequences normally within 100 kb of the coding region of a locus, but they may also be more distant from the coding region, which affect the expression of the gene (including transcription of the gene, and translation, splicing, stability or the like of the messenger RNA).

"Substantial homology or similarity". A nucleic acid or fragment thereof is "substantially homologous" ("or substantially similar") to another if, when optimally aligned (with appropriate nucleotide insertions or deletions) with the other nucleic acid (or its complementary strand), there is nucleotide sequence identity in at least about 60% of the nucleotide bases, usually at least about 70%, more usually at least about 80%, preferably at least about 90%, and more preferably at least about 95-98% of the nucleotide bases.

Alternatively, substantial homology or (similarity) exists when a nucleic acid or fragment thereof will hybridize to another nucleic acid (or a complementary strand thereof) under selective hybridization conditions, to a strand, or to its complement. Selectivity of hybridization exists when hybridization which is substantially more selective than total lack of specificity occurs. Typically, selective hybridization will occur when there is at least about 55% homology over a stretch of at least about 14 nucleotides, preferably at least about 65%, more preferably at least about 75%, and most preferably at least about 90%. See, Kanehisa, 1984. The length of homology comparison, as described, may be over longer stretches, and in certain embodiments will often be over a stretch of at least about nine nucleotides, usually at least about 20 nucleotides, more usually at least about 24 nucleotides, typically at least about 28 nucleotides, more typically at least about 32 nucleotides, and preferably at least about 36 or more nucleotides.

Nucleic acid hybridization will be affected by such conditions as salt concentration, temperature, or organic solvents, in addition to the base composition, length of the complementary strands, and the number of nucleotide base mismatches between the hybridizing nucleic acids, as will be readily appreciated by those skilled in the art. Stringent temperature conditions will generally include temperatures in excess of 30° C. typically in excess of 37° C. and preferably in excess of 45° C. Stringent salt conditions will ordinarily be less than 1000 mM, typically less than 500 mM, and preferably less than 200 mM. However, the combination of parameters is much more important than the measure of any single parameter. See, e.g., Wetmur & Davidson, 1968.

Probe sequences may also hybridize specifically to duplex DNA under certain conditions to form triplex or other higher order DNA complexes. The preparation of such probes and suitable hybridization conditions are well known in the art.

The terms "substantial homology" or "substantial identity", when referring to polypeptides, indicate that the polypeptide or protein in question exhibits at least about 30% identity with an entire naturally-occurring protein or a portion thereof, usually at least about 70% identity, and preferably at least about 95% identity.

"Substantially similar function" refers to the function of a modified nucleic acid or a modified protein, with reference

5,747,282

25

to the wild-type BRCA1 nucleic acid or wild-type BRCA1 polypeptide. The modified polypeptide will be substantially homologous to the wild-type BRCA1 polypeptide and will have substantially the same function. The modified polypeptide may have an altered amino acid sequence and/or may contain modified amino acids. In addition to the similarity of function, the modified polypeptide may have other useful properties, such as a longer half-life. The similarity of function (activity) of the modified polypeptide may be substantially the same as the activity of the wild-type BRCA1 polypeptide. Alternatively, the similarity of function (activity) of the modified polypeptide may be higher than the activity of the wild-type BRCA1 polypeptide. The modified polypeptide is synthesized using conventional techniques, or is encoded by a modified nucleic acid and produced using conventional techniques. The modified nucleic acid is prepared by conventional techniques. A nucleic acid with a function substantially similar to the wild-type BRCA1 gene function produces the modified protein described above.

Homology, for polypeptides, is typically measured using sequence analysis software. See, e.g., the Sequence Analysis Software Package of the Genetics Computer Group, University of Wisconsin Biotechnology Center, 910 University Avenue, Madison, Wisc. 53705. Protein analysis software matches similar sequences using measure of homology assigned to various substitutions, deletions and other modifications. Conservative substitutions typically include substitutions within the following groups: glycine, alanine; valine, isoleucine, leucine; aspartic acid, glutamic acid; asparagine, glutamine; serine, threonine; lysine, arginine; and phenylalanine, tyrosine.

A polypeptide "fragment," "portion" or "segment" is a stretch of amino acid residues of at least about five to seven contiguous amino acids, often at least about seven to nine contiguous amino acids, typically at least about nine to 13 contiguous amino acids and, most preferably, at least about 20 to 30 or more contiguous amino acids.

The polypeptides of the present invention, if soluble, may be coupled to a solid-phase support, e.g., nitrocellulose, nylon, column packing materials (e.g., Sepharose beads), magnetic beads, glass wool, plastic, metal, polymer gels, cells, or other substrates. Such supports may take the form, for example, of beads, wells, dipsticks, or membranes.

"Target region" refers to a region of the nucleic acid which is amplified and/or detected. The term "target sequence" refers to a sequence with which a probe or primer will form a stable hybrid under desired conditions.

The practice of the present invention employs, unless otherwise indicated, conventional techniques of chemistry, molecular biology, microbiology, recombinant DNA, genetics, and immunology. See, e.g., Maniatis et al., 1982; Sambrook et al., 1989; Ausubel et al., 1992; Glover, 1985; Anand, 1992; Guthrie & Fink, 1991. A general discussion of techniques and materials for human gene mapping, including mapping of human chromosome 17q, is provided, e.g., in White and Lalouel, 1988.

Preparation of Recombinant or Chemically Synthesized Nucleic Acids; Vectors, Transformation, Host Cells

Large amounts of the polynucleotides of the present invention may be produced by replication in a suitable host cell. Natural or synthetic polynucleotide fragments coding for a desired fragment will be incorporated into recombinant polynucleotide constructs, usually DNA constructs, capable of introduction into and replication in a prokaryotic or eukaryotic cell. Usually the polynucleotide constructs will

26

be suitable for replication in a unicellular host, such as yeast or bacteria, but may also be intended for introduction to (with and without integration within the genome) cultured mammalian or plant or other eukaryotic cell lines. The purification of nucleic acids produced by the methods of the present invention is described, e.g., in Sambrook et al., 1989 or Ausubel et al., 1992.

The polynucleotides of the present invention may also be produced by chemical synthesis, e.g., by the phosphoramidite method described by Beaucage & Carruthers, 1981 or the triester method according to Matteucci and Caruthers, 1981, and may be performed on commercial, automated oligonucleotide synthesizers. A double-stranded fragment may be obtained from the single-stranded product of chemical synthesis either by synthesizing the complementary strand and annealing the strands together under appropriate conditions or by adding the complementary strand using DNA polymerase with an appropriate primer sequence.

Polynucleotide constructs prepared for introduction into a prokaryotic or eukaryotic host may comprise a replication system recognized by the host, including the intended polynucleotide fragment encoding the desired polypeptide, and will preferably also include transcription and translational initiation regulatory sequences operably linked to the polypeptide encoding segment. Expression vectors may include, for example, an origin of replication or autonomously replicating sequence (ARS) and expression control sequences, a promoter, an enhancer and necessary processing information sites, such as ribosome-binding sites, RNA splice sites, polyadenylation sites, transcriptional terminator sequences, and mRNA stabilizing sequences. Secretion signals may also be included where appropriate, whether from a native BRCA1 protein or from other receptors or from secreted polypeptides of the same or related species, which allow the protein to cross and/or lodge in cell membranes, and thus attain its functional topology, or be secreted from the cell. Such vectors may be prepared by means of standard recombinant techniques well known in the art and discussed, for example, in Sambrook et al., 1989 or Ausubel et al. 1992.

An appropriate promoter and other necessary vector sequences will be selected so as to be functional in the host, and may include, when appropriate, those naturally associated with BRCA1 genes. Examples of workable combinations of cell lines and expression vectors are described in Sambrook et al., 1989 or Ausubel et al., 1992; see also, e.g., Metzger et al., 1988. Many useful vectors are known in the art and may be obtained from such vendors as Stratagene, New England Biolabs, Promega Biotech, and others. Promoters such as the trp, lac and phage promoters, tRNA promoters and glycolytic enzyme promoters may be used in prokaryotic hosts. Useful yeast promoters include promoter regions for metallothionein, 3-phosphoglycerate kinase or other glycolytic enzymes such as enolase or glyceraldehyde-3-phosphate dehydrogenase, enzymes responsible for maltose and galactose utilization, and others. Vectors and promoters suitable for use in yeast expression are further described in Hitzeman et al., EP 73,675A. Appropriate non-native mammalian promoters might include the early and late promoters from SV40 (Fiers et al., 1978) or promoters derived from murine Moloney leukemia virus, mouse tumor virus, avian sarcoma viruses, adenovirus II, bovine papilloma virus or polyoma. In addition, the construct may be joined to an amplifiable gene (e.g., DHFR) so that multiple copies of the gene may be made. For appropriate enhancer and other expression control sequences, see also *Enhancers and Eukaryotic Gene Expression*, Cold Spring Harbor Press, Cold Spring Harbor, N. Y. (1983).

A000177

5,747,282

27

While such expression vectors may replicate autonomously, they may also replicate by being inserted into the genome of the host cell, by methods well known in the art.

Expression and cloning vectors will likely contain a selectable marker, a gene encoding a protein necessary for survival or growth of a host cell transformed with the vector. The presence of this gene ensures growth of only those host cells which express the inserts. Typical selection genes encode proteins that a) confer resistance to antibiotics or other toxic substances, e.g. ampicillin, neomycin, methotrexate, etc.; b) complement auxotrophic deficiencies, or c) supply critical nutrients not available from complex media, e.g., the gene encoding D-alanine racemase for *Bacilli*. The choice of the proper selectable marker will depend on the host cell, and appropriate markers for different hosts are well known in the art.

The vectors containing the nucleic acids of interest can be transcribed *in vitro*, and the resulting RNA introduced into the host cell by well-known methods, e.g., by injection (see, Kubo et al., 1988), or the vectors can be introduced directly into host cells by methods well known in the art, which vary depending on the type of cellular host, including electroporation; transfection employing calcium chloride, rubidium chloride, calcium phosphate, DEAE-dextran, or other substances; microprojectile bombardment; lipofection; infection (where the vector is an infectious agent, such as a retroviral genome); and other methods. See generally, Sambrook et al., 1989 and Ausubel et al., 1992. The introduction of the polynucleotides into the host cell by any method known in the art, including, *inter alia*, those described above, will be referred to herein as "transformation." The cells into which have been introduced nucleic acids described above are meant to also include the progeny of such cells.

Large quantities of the nucleic acids and polypeptides of the present invention may be prepared by expressing the BRCA1 nucleic acids or portions thereof in vectors or other expression vehicles in compatible prokaryotic or eukaryotic host cells. The most commonly used prokaryotic hosts are strains of *Escherichia coli*, although other prokaryotes, such as *Bacillus subtilis* or *Pseudomonas* may also be used.

Mammalian or other eukaryotic host cells, such as those of yeast, filamentous fungi, plant, insect, or amphibian or avian species, may also be useful for production of the proteins of the present invention. Propagation of mammalian cells in culture is *per se* well known. See, Jakoby and Pastan, 1979. Examples of commonly used mammalian host cell lines are VERO and HeLa cells, Chinese hamster ovary (CHO) cells, and WI38, BHK, and COS cell lines, although it will be appreciated by the skilled practitioner that other cell lines may be appropriate, e.g., to provide higher expression, desirable glycosylation patterns, or other features.

Clones are selected by using markers depending on the mode of the vector construction. The marker may be on the same or a different DNA molecule, preferably the same DNA molecule. In prokaryotic hosts, the transformant may be selected, e.g., by resistance to ampicillin, tetracycline or other antibiotics. Production of a particular product based on temperature sensitivity may also serve as an appropriate marker.

Prokaryotic or eukaryotic cells transformed with the polynucleotides of the present invention will be useful not only for the production of the nucleic acids and polypeptides of the present invention, but also, for example, in studying the characteristics of BRCA1 polypeptides.

Antisense polynucleotide sequences are useful in preventing or diminishing the expression of the BRCA1 locus, as

28

will be appreciated by those skilled in the art. For example, polynucleotide vectors containing all or a portion of the BRCA1 locus or other sequences from the BRCA1 region (particularly those flanking the BRCA1 locus) may be placed under the control of a promoter in an antisense orientation and introduced into a cell. Expression of such an antisense construct within a cell will interfere with BRCA1 transcription and/or translation and/or replication.

The probes and primers based on the BRCA1 gene sequences disclosed herein are used to identify homologous BRCA1 gene sequences and proteins in other species. These BRCA1 gene sequences and proteins are used in the diagnostic/prognostic, therapeutic and drug screening methods described herein for the species from which they have been isolated.

Methods of Use: Nucleic Acid Diagnosis and Diagnostic Kits

In order to detect the presence of a BRCA1 allele predisposing an individual to cancer, a biological sample such as blood is prepared and analyzed for the presence or absence of susceptibility alleles of BRCA1. In order to detect the presence of neoplasia, the progression toward malignancy of a precursor lesion, or as a prognostic indicator, a biological sample of the lesion is prepared and analyzed for the presence or absence of mutant alleles of BRCA1. Results of these tests and interpretive information are returned to the health care provider for communication to the tested individual. Such diagnoses may be performed by diagnostic laboratories, or, alternatively, diagnostic kits are manufactured and sold to health care providers or to private individuals for self-diagnosis.

Initially, the screening method involves amplification of the relevant BRCA1 sequences. In another preferred embodiment of the invention, the screening method involves a non-PCR based strategy. Such screening methods include two-step label amplification methodologies that are well known in the art. Both PCR and non-PCR based screening strategies can detect target sequences with a high level of sensitivity.

The most popular method used today is target amplification. Here, the target nucleic acid sequence is amplified with polymerases. One particularly preferred method using polymerase-driven amplification is the polymerase chain reaction (PCR). The polymerase chain reaction and other polymerase-driven amplification assays can achieve over a million-fold increase in copy number through the use of polymerase-driven amplification cycles. Once amplified, the resulting nucleic acid can be sequenced or used as a substrate for DNA probes.

When the probes are used to detect the presence of the target sequences (for example, in screening for cancer susceptibility), the biological sample to be analyzed, such as blood or serum, may be treated, if desired, to extract the nucleic acids. The sample nucleic acid may be prepared in various ways to facilitate detection of the target sequence; e.g. denaturation, restriction digestion, electrophoresis or dot blotting. The targeted region of the analyte nucleic acid usually must be at least partially single-stranded to form hybrids with the targeting sequence of the probe. If the sequence is naturally single-stranded, denaturation will not be required. However, if the sequence is double-stranded, the sequence will probably need to be denatured. Denaturation can be carried out by various techniques known in the art.

Analyte nucleic acid and probe are incubated under conditions which promote stable hybrid formation of the target sequence in the probe with the putative targeted

30

which turns over a chemiluminescent substrate. For methods for labeling nucleic acid probes according to this embodiment see Martin et al., 1990. In a second example, the small ligand is recognized by a second ligand-enzyme conjugate that is capable of specifically complexing to the first ligand. A well known embodiment of this example is the biotin-avidin type of interactions. For methods for labeling nucleic acid probes and their use in biotin-avidin based assays see Rigby et al., 1977 and Nguyen et al., 1992.

It is also contemplated within the scope of this invention that the nucleic acid probe assays of this invention will employ a cocktail of nucleic acid probes capable of detecting BRCA1. Thus, in one example to detect the presence of BRCA1 in a cell sample, more than one probe complementary to BRCA1 is employed and in particular the number of different probes is alternatively 2, 3, or 5 different nucleic acid probe sequences. In another example, to detect the presence of mutations in the BRCA1 gene sequence in a patient, more than one probe complementary to BRCA1 is employed where the cocktail includes probes capable of binding to the allele-specific mutations identified in populations of patients with alterations in BRCA1. In this embodiment, any number of probes can be used, and will preferably include probes corresponding to the major gene mutations identified as predisposing an individual to breast cancer. Some candidate probes contemplated within the scope of the invention include probes that include the allele-specific mutations identified in Tables 11 and 12 and those that have the BRCA1 regions corresponding to SEQ ID NO:1 both 5' and 3' to the mutation site.

As noted above, non-PCR based screening assays are also contemplated in this invention. An exemplary non-PCR based procedure is provided in Example 11. This procedure hybridizes a nucleic acid probe (or an analog such as a methyl phosphonate backbone replacing the normal phosphodiester), to the low level DNA target. This probe may have an enzyme covalently linked to the probe, such that the covalent linkage does not interfere with the specificity of the hybridization. This enzyme-probe-conjugate-target nucleic acid complex can then be isolated away from the free probe enzyme conjugate and a substrate is added for enzyme detection. Enzymatic activity is observed as a change in color development or luminescent output resulting in a 10^3 - 10^6 increase in sensitivity. For an example relating to the preparation of oligodeoxynucleotide-alkaline phosphatase conjugates and their use as hybridization probes see Jablonski et al., 1986.

The neoplastic condition of lesions can also be detected on the basis of the alteration of wild-type BRCA1 polypeptide. Such alterations can be determined by sequence analysis in accordance with conventional techniques. More preferably, antibodies (polyclonal or monoclonal) are used to detect differences in, or the absence of BRCA1 peptides. The antibodies may be prepared as discussed above under the heading "Antibodies" and as further shown in Examples 12 and 13. Other techniques for raising and purifying antibodies are well known in the art and any such techniques may be chosen to achieve the preparations claimed in this invention. In a preferred embodiment of the invention, antibodies will immunoprecipitate BRCA1 proteins from solution as well as react with BRCA1 protein on Western or immunoblots of polyacrylamide gels. In another preferred embodiment, antibodies will detect BRCA1 proteins in paraffin or frozen tissue sections, using immunocytochemical techniques.

Two-step label amplification methodologies are known in the art. These assays work on the principle that a small ligand (such as digoxigenin, biotin, or the like) is attached to a nucleic acid probe capable of specifically binding BRCA1. Exemplary probes are provided in Table 9 of this patent application and additionally include the nucleic acid probe corresponding to nucleotide positions 3631 to 3930 of SEQ ID NO:1. Allele specific probes are also contemplated within the scope of this example and exemplary allele specific probes include probes encompassing the predisposing mutations summarized in Tables 11 and 12 of this patent application.

Preferred embodiments relating to methods for detecting BRCA1 or its mutations include enzyme linked immunosorbent assays (ELISA), radioimmunoassays (RIA), immunoradiometric assays (IRMA) and immunoenzymatic assays (IEMA), including sandwich assays using monoclonal and/or polyclonal antibodies. Exemplary sandwich assays are described by David et al. in U.S. Pat. Nos. 5,437,110 and 4,486,530, hereby incorporated by reference, and exemplified in Example 14.

Methods of Use: Drug Screening

In one example, the small ligand attached to the nucleic acid probe is specifically recognized by an antibody-enzyme conjugate. In one embodiment of this example, digoxigenin is attached to the nucleic acid probe. Hybridization is detected by an antibody-alkaline phosphatase conjugate

This invention is particularly useful for screening compounds by using the BRCA1 polypeptide or binding fragment thereof in any of a variety of drug screening techniques.

The BRCA1 polypeptide or fragment employed in such a test may either be free in solution, affixed to a solid support, or borne on a cell surface. One method of drug screening utilizes eucaryotic or procaryotic host cells which are stably

5,747,282

31

transformed with recombinant polynucleotides expressing the polypeptide or fragment, preferably in competitive binding assays. Such cells, either in viable or fixed form, can be used for standard binding assays. One may measure, for example, for the formation of complexes between a BRCA1 polypeptide or fragment and the agent being tested, or examine the degree to which the formation of a complex between a BRCA1 polypeptide or fragment and a known ligand is interfered with by the agent being tested.

Thus, the present invention provides methods of screening for drugs comprising contacting such an agent with a BRCA1 polypeptide or fragment thereof and assaying (i) for the presence of a complex between the agent and the BRCA1 polypeptide or fragment, or (ii) for the presence of a complex between the BRCA1 polypeptide or fragment and a ligand, by methods well known in the art. In such competitive binding assays the BRCA1 polypeptide or fragment is typically labeled. Free BRCA1 polypeptide or fragment is separated from that present in a protein:protein complex, and the amount of free (i.e., uncomplexed) label is a measure of the binding of the agent being tested to BRCA1 or its interference with BRCA1:ligand binding, respectively.

Another technique for drug screening provides high throughput screening for compounds having suitable binding affinity to the BRCA1 polypeptides and is described in detail in Geysen, PCT published application WO 84/03564, published on Sep. 13, 1984. Briefly stated, large numbers of different small peptide test compounds are synthesized on a solid substrate, such as plastic pins or some other surface. The peptide test compounds are reacted with BRCA1 polypeptide and washed. Bound BRCA1 polypeptide is then detected by methods well known in the art.

Purified BRCA1 can be coated directly onto plates for use in the aforementioned drug screening techniques. However, non-neutralizing antibodies to the polypeptide can be used to capture antibodies to immobilize the BRCA1 polypeptide on the solid phase.

This invention also contemplates the use of competitive drug screening assays in which neutralizing antibodies capable of specifically binding the BRCA1 polypeptide compete with a test compound for binding to the BRCA1 polypeptide or fragments thereof. In this manner, the antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants of the BRCA1 polypeptide.

A further technique for drug screening involves the use of host eukaryotic cell lines or cells (such as described above) which have a nonfunctional BRCA1 gene. These host cell lines or cells are defective at the BRCA1 polypeptide level. The host cell lines or cells are grown in the presence of drug compound. The rate of growth of the host cells is measured to determine if the compound is capable of regulating the growth of BRCA1 defective cells.

Methods of Use: Rational Drug Design

The goal of rational drug design is to produce structural analogs of biologically active polypeptides of interest or of small molecules with which they interact (e.g., agonists, antagonists, inhibitors) in order to fashion drugs which are, for example, more active or stable forms of the polypeptide, or which, e.g., enhance or interfere with the function of a polypeptide in vivo. See, e.g., Hodgson, 1991. In one approach, one first determines the three-dimensional structure of a protein of interest (e.g., BRCA1 polypeptide) or, for example, of the BRCA1-receptor or ligand complex, by x-ray crystallography, by computer modeling or most typically, by a combination of approaches. Less often, useful information regarding the structure of a polypeptide may be

32

gained by modeling based on the structure of homologous proteins. An example of rational drug design is the development of HIV protease inhibitors (Erickson et al., 1990). In addition, peptides (e.g., BRCA1 polypeptide) are analyzed by an alanine scan (Wells, 1991). In this technique, an amino acid residue is replaced by Ala, and its effect on the peptide's activity is determined. Each of the amino acid residues of the peptide is analyzed in this manner to determine the important regions of the peptide.

It is also possible to isolate a target-specific antibody, selected by a functional assay, and then to solve its crystal structure. In principle, this approach yields a pharmacore upon which subsequent drug design can be based. It is possible to bypass protein crystallography altogether by generating anti-idio-typic antibodies (anti-ids) to a functional, pharmacologically active antibody. As a mirror image of a mirror image, the binding site of the anti-ids would be expected to be an analog of the original receptor. The anti-id could then be used to identify and isolate peptides from banks of chemically or biologically produced banks of peptides. Selected peptides would then act as the pharmacore.

Thus, one may design drugs which have, e.g., improved BRCA1 polypeptide activity or stability or which act as inhibitors, agonists, antagonists, etc. of BRCA1 polypeptide activity. By virtue of the availability of cloned BRCA1 sequences, sufficient amounts of the BRCA1 polypeptide may be made available to perform such analytical studies as x-ray crystallography. In addition, the knowledge of the BRCA1 protein sequence provided herein will guide those employing computer modeling techniques in place of, or in addition to x-ray crystallography.

Methods of Use: Gene Therapy

According to the present invention, a method is also provided of supplying wild-type BRCA1 function to a cell which carries mutant BRCA1 alleles. Supplying such a function should suppress neoplastic growth of the recipient cells. The wild-type BRCA1 gene or a part of the gene may be introduced into the cell in a vector such that the gene remains extrachromosomal. In such a situation, the gene will be expressed by the cell from the extrachromosomal location. If a gene fragment is introduced and expressed in a cell carrying a mutant BRCA1 allele, the gene fragment should encode a part of the BRCA1 protein which is required for non-neoplastic growth of the cell. More preferred is the situation where the wild-type BRCA1 gene or a part thereof is introduced into the mutant cell in such a way that it recombines with the endogenous mutant BRCA1 gene present in the cell. Such recombination requires a double recombination event which results in the correction of the BRCA1 gene mutation. Vectors for introduction of genes both for recombination and for extrachromosomal maintenance are known in the art, and any suitable vector may be used. Methods for introducing DNA into cells such as electroporation, calcium phosphate co-precipitation and viral transduction are known in the art, and the choice of method is within the competence of the routineer. Cells transformed with the wild-type BRCA1 gene can be used as model systems to study cancer remission and drug treatments which promote such remission.

As generally discussed above, the BRCA1 gene or fragment, where applicable, may be employed in gene therapy methods in order to increase the amount of the expression products of such genes in cancer cells. Such gene therapy is particularly appropriate for use in both cancerous and pre-cancerous cells, in which the level of BRCA1 polypeptide is absent or diminished compared to normal

A000180

cells. It may also be useful to increase the level of expression of a given BRCA1 gene even in those tumor cells in which the mutant gene is expressed at a "normal" level, but the gene product is not fully functional.

Gene therapy would be carried out according to generally accepted methods, for example, as described by Friedman, 1991. Cells from a patient's tumor would be first analyzed by the diagnostic methods described above, to ascertain the production of BRCA1 polypeptide in the tumor cells. A virus or plasmid vector (see further details below), containing a copy of the BRCA1 gene linked to expression control elements and capable of replicating inside the tumor cells, is prepared. Suitable vectors are known, such as disclosed in U.S. Pat. No. 5,252,479 and PCT published application WO 93/07282. The vector is then injected into the patient, either locally at the site of the tumor or systemically (in order to reach any tumor cells that may have metastasized to other sites). If the transfected gene is not permanently incorporated into the genome of each of the targeted tumor cells, the treatment may have to be repeated periodically.

Gene transfer systems known in the art may be useful in the practice of the gene therapy methods of the present invention. These include viral and nonviral transfer methods. A number of viruses have been used as gene transfer vectors, including papovaviruses, e.g., SV40 (Madzak et al., 1992), adenovirus (Berkner, 1992; Berkner et al., 1988; Gorziglia and Kapikian, 1992; Quantin et al., 1992; Rosenfeld et al., 1992; Wilkinson et al., 1992; Stratford-Perricaudet et al., 1990), vaccinia virus (Moss, 1992), adenomatus-associated virus (Muzyczka, 1992; Ohi et al., 1990), herpesviruses including HSV and EBV (Margolskee, 1992; Johnson et al., 1992; Fink et al., 1992; Breakfield and Geller, 1987; Freese et al., 1990), and retroviruses of avian (Brandypadhyay and Temin, 1984; Petropoulos et al., 1992), murine (Miller, 1992; Miller et al., 1985; Sorge et al., 1984; Mann and Baltimore, 1985; Miller et al., 1988), and human origin (Shimada et al., 1991; Helseth et al., 1990; Page et al., 1990; Buchschacher and Panganiban, 1992). Most human gene therapy protocols have been based on disabled murine retroviruses.

Nonviral gene transfer methods known in the art include chemical techniques such as calcium phosphate coprecipitation (Graham and van der Eb, 1973; Pellicer et al., 1980); mechanical techniques, for example microinjection (Anderson et al., 1980; Gordon et al., 1980; Brinster et al., 1981; Constantini and Lacy, 1981); membrane fusion-mediated transfer via liposomes (Felgner et al., 1987; Wang and Huang, 1989; Kaneda et al., 1989; Stewart et al., 1992; Nabel et al., 1990; Lim et al., 1992); and direct DNA uptake and receptor-mediated DNA transfer (Wolff et al., 1990; Wu et al., 1991; Zenke et al., 1990; Wu et al., 1989b; Wolff et al., 1991; Wagner et al., 1990; Wagner et al., 1991; Cotten et al., 1990; Curiel et al., 1991a; Curiel et al., 1991b). Viral-mediated gene transfer can be combined with direct in vivo gene transfer using liposome delivery, allowing one to direct the viral vectors to the tumor cells and not into the surrounding nondividing cells. Alternatively, the retroviral vector producer cell line can be injected into tumors (Culver et al., 1992). Injection of producer cells would then provide a continuous source of vector particles. This technique has been approved for use in humans with inoperable brain tumors.

In an approach which combines biological and physical gene transfer methods, plasmid DNA of any size is combined with a polylysine-conjugated antibody specific to the adenovirus hexon protein, and the resulting complex is bound to an adenovirus vector. The trimolecular complex is

then used to infect cells. The adenovirus vector permits efficient binding, internalization, and degradation of the endosome before the coupled DNA is damaged.

Liposome/DNA complexes have been shown to be capable of mediating direct in vivo gene transfer. While in standard liposome preparations the gene transfer process is nonspecific, localized in vivo uptake and expression have been reported in tumor deposits, for example, following direct in situ administration (Nabel, 1992).

Gene transfer techniques which target DNA directly to breast and ovarian tissues, e.g., epithelial cells of the breast or ovaries, is preferred. Receptor-mediated gene transfer, for example, is accomplished by the conjugation of DNA (usually in the form of covalently closed supercoiled plasmid) to a protein ligand via polylysine. Ligands are chosen on the basis of the presence of the corresponding ligand receptors on the cell surface of the target cell/tissue type. One appropriate receptor/ligand pair may include the estrogen receptor and its ligand, estrogen (and estrogen analogues). These ligand-DNA conjugates can be injected directly into the blood if desired and are directed to the target tissue where receptor binding and internalization of the DNA-protein complex occurs. To overcome the problem of intracellular destruction of DNA, coinfection with adenovirus can be included to disrupt endosome function.

The therapy involves two steps which can be performed singly or jointly. In the first step, prepubescent females who carry a BRCA1 susceptibility allele are treated with a gene delivery vehicle such that some or all of their mammary ductal epithelial precursor cells receive at least one additional copy of a functional normal BRCA1 allele. In this step, the treated individuals have reduced risk of breast cancer to the extent that the effect of the susceptible allele has been countered by the presence of the normal allele. In the second step of a preventive therapy, predisposed young females, in particular women who have received the proposed gene therapeutic treatment, undergo hormonal therapy to mimic the effects on the breast of a full term pregnancy.

Methods of Use: Peptide Therapy
Peptides which have BRCA1 activity can be supplied to cells which carry mutant or missing BRCA1 alleles. The sequence of the BRCA1 protein is disclosed (SEQ ID NO:2). Protein can be produced by expression of the cDNA sequence in bacteria, for example, using known expression vectors. Alternatively, BRCA1 polypeptide can be extracted from BRCA1-producing mammalian cells. In addition, the techniques of synthetic chemistry can be employed to synthesize BRCA1 protein. Any of such techniques can provide the preparation of the present invention which comprises the BRCA1 protein. The preparation is substantially free of other human proteins. This is most readily accomplished by synthesis in a microorganism or in vitro.

Active BRCA1 molecules can be introduced into cells by microinjection or by use of liposomes, for example. Alternatively, some active molecules may be taken up by cells, actively or by diffusion. Extracellular application of the BRCA1 gene product may be sufficient to affect tumor growth. Supply of molecules with BRCA1 activity should lead to partial reversal of the neoplastic state. Other molecules with BRCA1 activity (for example, peptides, drugs or organic compounds) may also be used to effect such a reversal. Modified polypeptides having substantially similar function are also used for peptide therapy.

Methods of Use: Transformed Hosts

Similarly, cells and animals which carry a mutant BRCA1 allele can be used as model systems to study and test for substances which have potential as therapeutic agents. The

cells are typically cultured epithelial cells. These may be isolated from individuals with BRCA1 mutations, either somatic or germline. Alternatively, the cell line can be engineered to carry the mutation in the BRCA1 allele, as described above. After a test substance is applied to the cells, the neoplastically transformed phenotype of the cell is determined. Any trait of neoplastically transformed cells can be assessed, including anchorage-independent growth, tumorigenicity in nude mice, invasiveness of cells, and growth factor dependence. Assays for each of these traits are known in the art.

Animals for testing therapeutic agents can be selected after mutagenesis of whole animals or after treatment of germ-line cells or zygotes. Such treatments include insertion of mutant BRCA1 alleles, usually from a second animal species, as well as insertion of disrupted homologous genes. Alternatively, the endogenous BRCA1 gene(s) of the animals may be disrupted by insertion or deletion mutation or other genetic alterations using conventional techniques (Capecci, 1989; Valancius and Smithies, 1991; Hasty et al., 1991; Shinkai et al., 1992; Mombaerts et al., 1992; Philpott et al., 1992; Snouwaert et al., 1992; Donehower et al., 1992). After test substances have been administered to the animals, the growth of tumors must be assessed. If the test substance prevents or suppresses the growth of tumors, then the test substance is a candidate therapeutic agent for the treatment of the cancers identified herein. These animal models provide an extremely important testing vehicle for potential therapeutic products.

The present invention is described by reference to the following Examples, which are offered by way of illustration and are not intended to limit the invention in any manner. Standard techniques well known in the art or the techniques specifically described below were utilized.

EXAMPLE 1

Ascertain and Study Kindreds Likely to Have a 17q-Linked Breast Cancer Susceptibility Locus

Extensive cancer prone kindreds were ascertained by our University of Utah collaborators from a defined population providing a large set of extended kindreds with multiple cases of breast cancer and many relatives available to study. The large number of meioses present in these large kindreds provided the power to detect whether the BRCA1 locus was segregating, and increased the opportunity for informative

recombinants to occur within the small region being investigated. This vastly improved the chances of establishing linkage to the BRCA1 region, and greatly facilitated the reduction of the BRCA1 region to a manageable size, which permits identification of the BRCA1 locus itself.

Each kindred was extended through all available connecting relatives by our collaborators, and to all informative first degree relatives of each proband or cancer case. For these kindreds, additional breast cancer cases and individuals with cancer at other sites of interest (e.g. ovarian) who also appeared in the kindreds were identified through the tumor registry linked files. All breast cancers reported in the kindred which were not confirmed in the Utah Cancer Registry were researched. Medical records or death certificates were obtained for confirmation of all cancers. Each key connecting individual and all informative individuals were invited by our collaborators to participate by providing a blood sample from which DNA was extracted. They also sampled spouses and relatives of deceased cases so that the genotype of the deceased cases could be inferred from the genotypes of their relatives.

Ten kindreds which had three or more cancer cases with inferable genotypes were selected for linkage studies to 17q markers from a set of 29 kindreds originally ascertained for a study of proliferative breast disease and breast cancer (Skolnick et al., 1990). The criterion for selection of these kindreds was the presence of two sisters or a mother and her daughter with breast cancer. Additionally, two kindreds which have been studied by our collaborators since 1980 as part of their breast cancer linkage studies (K1001, K9018), six kindreds ascertained for the presence of clusters of breast and/or ovarian cancer (K2019, K2073, K2079, K2080, K2039, K2082) and a self-referred kindred with early onset breast cancer (K2035) were included. These kindreds were investigated and expanded in our collaborators clinic in the manner described above. Table 1 displays the characteristics of these 19 kindreds which are the subject of subsequent examples. In Table 1, for each kindred the total number of individuals in our database, the number of typed individuals, and the minimum, median, and maximum age at diagnosis of breast/ovarian cancer are reported. Kindreds are sorted in ascending order of median age at diagnosis of breast cancer. Four women diagnosed with both ovarian and breast cancer are counted in both categories.

TABLE 1

KINDRED	Description of the 19 Kindreds									
	No. of		Breast				Ovarian			
	Individuals		Age at Dx				Age at Dx			
	Total	Sample	# Aff.	Min.	Med.	Max.	# Aff.	Min.	Med.	Max.
1910	15	10	4	27	34	49	—	—	—	—
1001	133	98	13	28	37	64	—	—	—	—
2035	42	25	8	28	37	45	1	—	60	—
2027	21	11	4	34	38	41	—	—	—	—
9018	54	17	9	30	40	72	2	46	48	50
1925	50	27	4	39	42	53	—	—	—	—
1927	49	29	5	32	42	51	—	—	—	—
1911	28	21	7	28	42	76	—	—	—	—
1929	16	11	4	34	43	73	—	—	—	—
1901	35	19	10	31	44	76	—	—	—	—
2082	180	105	20	27	47	67	10	45	52	66
2019	42	19	10	42	53	79	—	—	—	—
1900	70	23	8	45	55	70	1	—	78	—

TABLE 1-continued

<u>Description of the 19 Kindreds</u>										
KINDRED	No. of		<u>Breast</u>				<u>Ovarian</u>			
	<u>Individuals</u>		# Aff.	<u>Age at Dx</u>			# Aff.	<u>Age at Dx</u>		
	Total	Sample		Min.	Med.	Max.		Min.	Med.	Max.
2080	264	74	22+	27	55	92	4	45	53	71
2073	57	29	9	35	57	80	—	—	—	—
1917	16	6	4	43	58	61	—	—	—	—
1920	22	14	3	62	63	68	—	—	—	—
2079	136	18	14	38	66	84	4	52	59	65
2039	87	40	14	44	68	88	4	41	51	75

+Includes one case of male breast cancer.

EXAMPLE 2

Selection of Kindreds Which are Linked to Chromosome 17q and Localization of BRCA1 to the Interval Mfd15-Mfd188

For each sample collected in these 19 kindreds, DNA was extracted from blood (or in two cases from paraffin-embedded tissue blocks) using standard laboratory protocols. Genotyping in this study was restricted to short tandem repeat (STR) markers since, in general, they have high heterozygosity and PCR methods offer rapid turnaround while using very small amounts of DNA. To aid in this effort, four such STR markers on chromosome 17 were developed by screening a chromosome specific cosmid library for CA positive clones. Three of these markers localized to the long arm: (46E6, Easton et al., 1993); (42D6, Easton et al., 1993); 26C2 (D17S514, Oliphant et al., 1991), while the other, 12G6 (D17S513, Oliphant et al., 1991), localized to the short arm near the p53 tumor suppressor locus. Two of these, 42D6 and 46E6, were submitted to the Breast Cancer Linkage Consortium for typing of breast cancer families by investigators worldwide. Oligonucleotide sequences for markers not developed in our laboratory were obtained from published reports, or as part of the Breast Cancer Linkage Consortium, or from other investigators. All genotyping films were scored blindly with a standard lane marker used to maintain consistent coding of alleles. Key samples in the four kindreds presented here underwent duplicate typing for all relevant markers. All 19 kindreds have been typed for two polymorphic CA repeat markers: 42D6 (D17S588), a CA repeat isolated in our laboratory, and Mfd15 (D17S250), a CA repeat provided by J. Weber (Weber et al., 1990). Several sources of probes were used to create genetic markers on chromosome 17, specifically chromosome 17 cosmid and lambda phage libraries created from sorted chromosomes by the Los Alamos National Laboratories (van Dilla et al., 1986).

LOD scores for each kindred with these two markers (42D6, Mfd15) and a third marker, Mfd188 (D17S579, Hall et al., 1992), located roughly midway between these two markers, were calculated for two values of the recombination fraction, 0.001 and 0.1. (For calculation of LOD scores, see Oh, 1985). Likelihoods were computed under the model derived by Claus et al., 1991, which assumes an estimated gene frequency of 0.003, a lifetime risk in gene carriers of about 0.80, and population based age-specific risks for breast cancer in non-gene carriers. Allele frequencies for the three markers used for the LOD score calculations were calculated from our own laboratory typings of unrelated individuals in the CEPH panel (White and Lalouel, 1988).

Table 2 shows the results of the pairwise linkage analysis of each kindred with the three markers 42D6, Mfd188, and Mfd15.

TABLE 2

Pairwise Linkage Analysis of Kindreds						
KINDRED	Mfd15 (D17S250)		Mfd188 (D17S579)		42D6 (D17S588)	
	Recombination		Recombination		Recombination	
	0.001	0.1	0.001	0.1	0.001	0.1
1910	0.06	0.30	0.06	0.30	0.06	0.30
1001	-0.30	-0.09	NT	NT	-0.52	-0.19
2035	2.34	1.85	0.94	0.90	2.34	1.82
2027	-1.22	-0.33	-1.20	-0.42	-1.16	-0.33
9018	-0.54	-0.22	-0.17	-0.10	0.11	0.07
1925	1.08	0.79	0.55	0.38	-0.11	-0.07
1927	-0.41	0.01	-0.35	0.07	-0.44	-0.02
1911	-0.27	-0.13	-0.43	-0.23	0.49	0.38
1929	-0.49	-0.25	NT	NT	-0.49	-0.25
1901	1.50	1.17	0.78	0.57	0.65	0.37
2082	4.25	3.36	6.07	5.11	2.00	3.56
2019	-0.10	-0.01	-0.11	-0.05	-0.18	-0.10
1900	-0.14	-0.11	NT	NT	-0.12	-0.05
2080	-0.16	-0.04	0.76	0.74	-1.25	-0.58
2073	-0.41	-0.29	0.63	0.49	-0.23	-0.13
1917	-0.02	-0.02	NT	NT	-0.01	0.00
1920	-0.03	-0.02	NT	NT	0.00	0.00
2079	0.02	0.01	-0.01	-0.01	0.01	0.01
2039	-1.67	-0.83	0.12	0.59	-1.15	0.02

NT—Kindred not typed for mfd188.

Using a criterion for linkage to 17q of a LOD score>1.0 for at least one locus under the CASH model (Claus et al., 1991), four of the 19 kindreds appeared to be linked to 17q (K1901, K1925, K2035, K2082). A number of additional kindreds showed some evidence of linkage but at this time could not be definitively assigned to the linked category. These included kindreds K1911, K2073, K2039, and K2080. Three of the 17q-linked kindreds had informative recombinants in this region and these are detailed below.

Kindred 2082 is the largest 17q-linked breast cancer family reported to date by any group. The kindred contains 20 cases of breast cancer, and ten cases of ovarian cancer. Two cases have both ovarian and breast cancer. The evidence of linkage to 17q for this family is overwhelming; the LOD score with the linked haplotype is over 6.0, despite the existence of three cases of breast cancer which appear to be sporadic, i.e., these cases share no part of the linked haplotype between Mfd15 and 42D6. These three sporadic cases were diagnosed with breast cancer at ages 46, 47, and 54. In smaller kindreds, sporadic cancers of this type greatly confound the analysis of linkage and the correct identification of

key recombinants. The key recombinant in the 2082 kindred is a woman who developed ovarian cancer at age 45 whose mother and aunt had ovarian cancer at ages 58 and 66, respectively. She inherited the linked portion of the haplotype for both Mfd188 and 42D6 while inheriting unlinked alleles at Mfd15; this recombinant event placed BRCA1 distal to Mfd15.

K1901 is typical of early-onset breast cancer kindreds. The kindred contains 10 cases of breast cancer with a median age at diagnosis of 43.5 years of age; four cases were diagnosed under age 40. The LOD score for this kindred with the marker 42D6 is 1.5, resulting in a posterior probability of 17q-linkage of 0.96. Examination of haplotypes in this kindred identified a recombinant haplotype in an obligate male carrier and his affected daughter who was diagnosed with breast cancer at age 45. Their linked allele for marker Mfd15 differs from that found in all other cases in the kindred (except one case which could not be completely inferred from her children). The two haplotypes are identical for Mfd188 and 42D6. Accordingly, data from Kindred 1901 would also place the BRCA1 locus distal to Mfd15.

Kindred 2035 is similar to K1901 in disease phenotype. The median age of diagnosis for the eight cases of breast cancer in this kindred is 37. One case also had ovarian cancer at age 60. The breast cancer cases in this family descend from two sisters who were both unaffected with breast cancer until their death in the eighth decade. Each branch contains four cases of breast cancer with at least one case in each branch having markedly early onset. This kindred has a LOD score of 2.34 with Mfd5. The haplotypes segregating with breast cancer in the two branches share an identical allele at Mfd15 but differ for the distal loci Mfd188 and NM23 (a marker typed as part of the consortium which is located just distal to 42D6 (Hall et al., 1992)). Although the two haplotypes are concordant for marker 42D6, it is likely that the alleles are shared identical by state (the same allele but derived from different ancestors), rather than identical by descent (derived from a common ancestor) since the shared allele is the second most common allele observed at this locus. By contrast the linked allele shared at Mfd15 has a frequency of 0.04. This is a key recombinant in our dataset as it is the sole recombinant in which BRCA1 segregated with the proximal portion of the haplotype, thus setting the distal boundary to the BRCA1 region. For this event not to be a key recombinant requires that a second mutant BRCA1 gene be present in a spouse marrying into the kindred who also shares the rare Mfd15 allele segregating with breast cancer in both branches of the kindred. This event has a probability of less than one in a thousand. The evidence from this kindred therefore placed the BRCA1 locus proximal to Mfd188.

EXAMPLE 3

Creation of a Fine Structure Map and Refinement of the BRCA1 Region to Mfd191-Mfd188 using Additional STR Polymorphisms

In order to improve the characterization of our recombinants and define closer flanking markers, a dense map of this relatively small region on chromosome 17q was required. The chromosome 17 workshop has produced a consensus map of this region (FIG. 1) based on a combination of genetic and physical mapping studies (Fain, 1992). This map contains both highly polymorphic STR polymorphisms, and a number of nonpolymorphic expressed genes. Because this map did not give details on the evidence for this order nor give any measure of local support for inversions in the order of adjacent loci, we viewed it as a rough guide for obtaining resources to be used for the development of new markers and construction of our own detailed genetic and physical map of a small region containing BRCA1. Our approach was to analyze existing STR markers provided by other investigators and any newly developed markers from our laboratory with respect to both a panel of meiotic (genetic) breakpoints identified using DNA from the CEPH reference families and a panel of somatic cell hybrids (physical breakpoints) constructed for this region. These markers included 26C2 developed in our laboratory which maps proximal to Mfd15, Mfd191 (provided by James Weber), THRA1 (Futreal et al., 1992a), and three polymorphisms kindly provided to us by Dr. Donald Black, NM23 (Hall et al. 1992), SCG40 (D17S181), and 6C1 (D17S293).

Genetic localization of markers. In order to localize new markers genetically within the region of interest, we have identified a number of key meiotic breakpoints within the region, both in the CEPH reference panel and in our large breast cancer kindred (K2082). Given the small genetic distance in this region, they are likely to be only a relatively small set of recombinants which can be used for this purpose, and they are likely to group markers into sets. The orders of the markers within each set can only be determined by physical mapping. However the number of genotypings necessary to position a new marker is minimized. These breakpoints are illustrated in Tables 3 and 4. Using this approach we were able to genetically order the markers THRA1, 6C1, SCG40, and Mfd191. As can be seen from Tables 3 and 4, THRA1 and MFD191 both map inside the Mfd15-Mfd188 region we had previously identified as containing the BRCA1 locus. In Tables 3 and 4, M/P indicates a maternal or paternal recombinant. A "1" indicates inherited allele is of grandpaternal origin, while a "0" indicates grandmaternal origin, and "-" indicates that the locus was untyped or uninformative.

TABLE 3

CEPH Recombinants									
Family	ID	M/P	Mfd15	THRA1	Mfd191	Mfd188	SCG40	6C1	42D6
13292	4	M	1	1	1	0	0	0	0
13294	4	M	1	1	1	0	0	0	0
13294	6	M	0	0	1	1	—	—	—
1334	3	M	1	1	1	1	1	0	0
1333	4	M	1	1	1	0	—	—	0
1333	6	M	0	0	1	1	—	—	1
1333	8	P	1	0	0	0	—	—	0
1377	8	M	0	—	0	0	0	0	1

5,747,282

41

TABLE 4

Kindred 2082 Recombinants							
Family	ID	M/P	Mfd15	Mfd191	Mfd188	SCG40	6C1 42D6
75		M	0	1	1	1	—
63		M	0	0	1	1	1
125		M	1	1	1	0	0
40		M	1	1	0	0	0

Analysis of markers Mfd15, Mfd188, Mfd191, and THRA1 in our recombinant families. Mfd15, Mfd188, Mfd191 and THRA1 were typed in our recombinant families and examined for additional information to localize the BRCA1 locus. In kindred 1901, the Mfd15 recombinant was recombinant for THRA1 but uninformative for Mfd191, thus placing BRCA1 distal to THRA1. In K2082, the recombinant with Mfd15 also was recombinant with Mfd191, thus placing the BRCA1 locus distal to Mfd191 (Goldgar et al., 1994). Examination of THRA1 and Mfd191 in kindred K2035 yielded no further localization information as the two branches were concordant for both markers. However, SCG40 and 6C1 both displayed the same pattern as Mfd188, thus increasing our confidence in the localization information provided by the Mfd188 recombinant in this family. The BRCA1 locus, or at least a portion of it, therefore lies within an interval bounded by Mfd191 on the proximal side and Mfd188 on the distal side.

EXAMPLE 4

Development of Genetic and Physical Resources in the Region of Interest

To increase the number of highly polymorphic loci in the Mfd191-Mfd188 region, we developed a number of STR markers in our laboratory from cosmids and YACs which physically map to the region. These markers allowed us to further refine the region.

STSS were identified from genes known to be in the desired region to identify YACs which contained these loci, which were then used to identify subclones in cosmids, PIs or BACs. These subclones were then screened for the presence of a CA tandem repeat using a (CA)_n oligonucleotide (Pharmacia). Clones with a strong signal were selected preferentially, since they were more likely to represent CA-repeats which have a large number of repeats and/or are of near-perfect fidelity to the (CA)_n pattern. Both of these characteristics are known to increase the probability of polymorphism (Weber, 1990). These clones were sequenced directly from the vector to locate the repeat. We obtained a unique sequence on one side of the CA-repeat by using one of a set of possible primers complementary to the end of a CA-repeat, such as (GT)₁₀T. Based on this unique sequence, a primer was made to sequence back across the repeat in the other direction, yielding a unique sequence for design of a second primer flanking the CA-repeat. STRs were then screened for polymorphism on a small group of unrelated

42

individuals and tested against the hybrid panel to confirm their physical localization. New markers which satisfied these criteria were then typed in a set of 40 unrelated individuals from the Utah and CEPH families to obtain allele frequencies appropriate for the study population. Many of the other markers reported in this study were tested in a smaller group of CEPH unrelated individuals to obtain similarly appropriate allele frequencies.

Using the procedure described above, a total of eight polymorphic STRs was found from these YACS. Of the loci identified in this manner, four were both polymorphic and localized to the BRCA1 region. Four markers did not localize to chromosome 17, reflecting the chimeric nature of the YACs used. The four markers which were in the region were denoted AA1, ED2, 4-7, and YM29. AA1 and ED2 were developed from YACs positive for the RNU2 gene, 4-7 from an EPB3 YAC and YM29 from a cosmid which localized to the region by the hybrid panel. A description of the number of alleles, heterozygosity and source of these four and all other STR polymorphisms analyzed in the breast cancer kindreds is given below in Table 5.

TABLE 5

Polymorphic Short Tandem Repeat Markers Used for Fine Structure Mapping of the BRCA1 Locus							
Allele* Frequency (%)							
Clone	Gene	Na**	Heterozygosity	1	2	3	4 5 6
Mfd15	D17S250	10	0.82	26	22	15	7 7 23
THRA1	THRA1	5					
Mfd191	D17S776	7	0.55	48	20	11	7 7 7
ED2	D17S1327	12	0.55	62	9	8	5 5 11
AA1	D17S1326	7	0.83	28	28	25	8 6 5
CA375	D17S184	10	0.75	26	15	11	9 9 20
4-7	D17S1183	9	0.50	63	15	8	6 4 4
YM29	—	9	0.62	42	24	12	7 7 8
Mfd188	D17S579	12	0.92	33	18	8	8 8 25
SCG40	D17S181	14	0.90	20	18	18	10 8 35
42D6	D17S588	11	0.86	21	17	11	10 9 32
6C1	D17S293	7	0.75	30	30	11	11 9 9
Z109	D17S750	9	0.70	33	27	7	7 7 19
tdj1475	D17S1321	13	0.84	21	16	11	11 8 33
CF4	D17S1320	6	0.63	50	27	9	7 4 3
tdj1239	D17S1328	10	0.80	86	10	9	7 4 14
U5	D17S1325	13	0.83	19	16	12	10 9 34

*Allele codes 1-5 are listed in decreasing frequency; allele numbers do not correspond to fragment sizes. Allele 6 frequency is the joint frequency of all other alleles for each locus.
**Number of alleles seen in the genetically independent DNA samples used for calculating allele frequencies.

The four STR polymorphisms which mapped physically to the region (4-7, ED2, AA1, YM29) were analyzed in the meiotic, breakpoint panel shown initially in Tables 3 and 4. Tables 6 and 7 contain the relevant CEPH data and Kindred 2082 data for localization of these four markers. In the tables, M/P indicates a maternal or paternal recombinant. A "1" indicates inherited allele is of grandpaternal origin, while a "0" indicates grandmaternal origin, and "-" indicates that the locus was untyped or uninformative.

TABLE 6

Key Recombinants Used for Genetic Ordering of New STR Loci Developed in Our Laboratory Within the BRCA1 Region of 17q													
CEPH Family	ID	M/P	Mfd15	THRA1	MFD191	ED2	AA1	Z109	4-7	YM29	Mfd188	SCG40	42D6
13292	4	M	1	1	1	1	1	0	0	0	0	0	0
13294	4	M	1	0	0	—	0	—	—	—	0	—	—
13294	6	M	0	0	1	—	1	—	—	—	1	—	—
1333	4	M	1	1	1	—	0	—	—	0	0	—	0
1333	6	M	0	0	1	—	1	—	—	1	1	—	1
1333	3	M	0	0	1	—	—	—	1	1	1	—	1

TABLE 7

Kindred 2082 Recombinants										
ID	M/P	Mfd15	Mfd191	ED2	AA1	4-7	YM29	Mfd188	SCG40	42D6
63	M	0	0	1	—	1	1	1	1	1
125	M	1	1	1	—	1	1	1	0	0
40	M	1	1	0	—	0	—	0	0	0
22	P	0	0	1	1	1	1	1	1	1

From CEPH 1333-04, we see that AA1 and YM29 must lie distal to Mfd191. From 13292, it can be inferred that both AA1 and ED2 are proximal to 4-7, YM29, and Mfd188. The recombinants found in K2082 provide some additional ordering information. Three independent observations (individual numbers 22, 40, & 63) place AA1, ED2, 4-7, and YM29, and Mfd188 distal to Mfd191, while ID 125 places 4-7, YM29, and Mfd188 proximal to SCG40. No genetic information on the relative ordering within the two clusters of markers AA1/ED2 and 4-7/YM29/Mfd188 was obtained from the genetic recombinant analysis. Although ordering loci with respect to hybrids which are known to contain “holes” in which small pieces of interstitial human DNA may be missing is problematic, the hybrid patterns indicate that 4-7 lies above both YM29 and Mfd188.

EXAMPLE 5

Genetic Analyses of Breast Cancer Kindreds with Markers AA1,4-7, ED2, and YM29

In addition to the three kindreds containing key recombinants which have been discussed previously, kindred

K2039 was shown through analysis of the newly developed STR markers to be linked to the region and to contain a useful recombinant.

Table 8 defines the haplotypes (shown in coded form) of the kindreds in terms of specific marker alleles at each locus and their respective frequencies. In Table 8, alleles are listed in descending order of frequency; frequencies of alleles 1-5 for each locus are given in Table 5. Haplotypes coded H are BRCA1 associated haplotypes, P designates a partial H haplotype, and an R indicates an observable recombinant haplotype. As evident in Table 8, not all kindreds were typed for all markers; moreover, not all individuals within a kindred were typed for an identical set of markers, especially in K2082. With one exception, only haplotypes inherited from affected or at-risk kindred members are shown; haplotypes from spouses marrying into the kindred are not described. Thus in a given sibship, the appearance of haplotypes X and Y indicates that both haplotypes from the affected/at-risk individual were seen and neither was a breast cancer associated haplotype.

TABLE 8

Breast Cancer Linked Haplotypes Found in the Three Kindreds															
Kin.	HAP	Mfd15	THRA1	Mfd191	tdj1475	ED2	AA1	Z109	CA375	4-7	YM29	Mfd188	SCG40	6C1	42D6
1901	H1	1	5	5	3	1	4	NI	NI	1	1	3	NI	NI	1
	R2	9	2	5	6	1	4	NI	NI	1	1	3	NI	NI	1
2082	H1	3	NI	4	6	6	1	NI	NI	2	1	4	2	NI	1
	P1	3	NI	4	NI	NI	NI	NI	NI	NI	NI	4	2	NI	1
	P2	3	NI	NI	NI	NI	NI	NI	NI	NI	NI	4	NI	NI	NI
	R1	6	NI	1	5	6	1	NI	NI	2	1	4	2	NI	1
	R2	6	NI	4	6	6	1	NI	NI	2	1	4	2	NI	1
	R3	3	NI	4	NI	6	1	NI	NI	2	1	4	1	NI	7
	R4	7	NI	1	NI	1	5	NI	NI	4	6	1	2	NI	1
	R5	3	NI	4	NI	NI	NI	NI	NI	NI	2	1	NI	NI	NI
	R6	3	NI	4	3	1	2	NI	NI	1	2	2	6	NI	6

TABLE 8-continued

Breast Cancer Linked Haplotypes Found in the Three Kindreds															
Kin.	HAP	Mfd15	THRA1	Mfd191	tdj1475	ED2	AA1	Z109	CA375	4-7	YM29	Mfd188	SCG40	6C1	42D6
2035	R7	3	NI	4	3	7	1	NI	NI	1	1	3	7	NI	4
	H1	8	2	1	NI	5	1	1	4	3	1	6	8	2	4
	H2	8	2	1	NI	5	1	1	2	1	1	2	3	1	4
	R2	8	2	1	NI	5	1	1	2	1	1	2	3	6	1

In kindred K1901, the new markers showed no observable recombination with breast cancer susceptibility, indicating that the recombination event in this kindred most likely took place between THRA1 and ED2. Thus, no new BRCA1 localization information was obtained based upon studying the four new markers in this kindred. In kindred 2082 the key recombinant individual has inherited the linked alleles for ED2, 4-7, AA1, and YM29, and was recombinant for tdj1474 indicating that the recombination event occurred in this individual between tdj1474 and ED2/AA1.

There are three haplotypes of interest in kindred K2035, H1, H2, and R2 shown in Table 8. H1 is present in the four cases and one obligate male carrier descendant from individual 17 while H2 is present or inferred in two cases and two obligate male carriers in descendants of individual 10. R2 is identical to H2 for loci between and including Mfd15 and SCG40, but has recombined between SCG40 and 42D6. Since we have established that BRCA1 is proximal to 42D6, this H2/R2 difference adds no further localization information. H1 and R2 share an identical allele at Mfd15, THRA1, AA1, and ED2 but differ for loci presumed distal to ED2, i.e., 4-7, Mfd188, SCG40, and 6C1. Although the two haplotypes are concordant for the 5th allele for marker YM29, a marker which maps physically between 4-7 and Mfd 188, it is likely that the alleles are shared identical by state rather than identical by descent since this allele is the most common allele at this locus with a frequency estimated in CEPH parents of 0.42. By contrast, the linked alleles shared at the Mfd15 and ED2 loci have frequencies of 0.04 and 0.09, respectively. They also share more common alleles at Mfd191 (frequency=0.52), THRA1, and AA1 (frequency=0.28). This is the key recombinant in the set as it is the sole recombinant in which breast cancer segregated with the proximal portion of the haplotype, thus setting the distal boundary. The evidence from this kindred therefore places the BRCA1 locus proximal to 4-7.

The recombination event in kindred 2082 which places BRCA1 distal to tdj1474 is the only one of the four events described which can be directly inferred; that is, the affected mother's genotype can be inferred from her spouse and offspring, and the recombinant haplotype can be seen in her affected daughter. In this family the odds in favor of affected individuals carrying BRCA1 susceptibility alleles are extremely high; the only possible interpretations of the data are that BRCA1 is distal to Mfd191 or alternatively that the purported recombinant is a sporadic case of ovarian cancer at age 44. Rather than a directly observable or inferred recombinant, interpretation of kindred 2035 depends on the observation of distinct 17q-haplotypes segregating in different and sometimes distantly related branches of the kindred. The observation that portions of these haplotypes have alleles in common for some markers while they differ at other markers places the BRCA1 locus in the shared region. The confidence in this placement depends on several factors: the relationship between the individuals carrying the respective haplotypes, the frequency of the shared allele, the

certainty with which the haplotypes can be shown to segregate with the BRCA1 locus, and the density of the markers in the region which define the haplotype. In the case of kindred 2035, the two branches are closely related, and each branch has a number of early onset cases which carry the respective haplotype. While two of the shared alleles are common, (Mfd191, THRA1), the estimated frequencies of the shared alleles at Mfd15, AA1, and ED2 are 0.04, 0.28, and 0.09, respectively. It is therefore highly likely that these alleles are identical by descent (derived from a common ancestor) rather than identical by state (the same allele but derived from the general population).

EXAMPLE 6

Refined Physical Mapping Studies Place the BRCA1 Gene in a Region Flanked by tdj1474 and USR

Since its initial localization to chromosome 17q in 1990 (Hall et al., 1990) a great deal of effort has gone into localizing the BRCA1 gene to a region small enough to allow implementation of effective positional cloning strategies to isolate the gene. The BRCA1 locus was first localized to the interval Mfd15 (D17S250)-42D6 (D17S588) by multipoint linkage analysis (Easton et al., 1993) in the collaborative Breast Cancer Linkage Consortium dataset consisting of 214 families collected worldwide. Subsequent refinements of the localization have been based upon individual recombinant events in specific families. The region THRA1-D17S183 was defined by Bowcock et al., 1993; and the region THRA1-D17S78 was defined by Simard et al., 1993.

We further showed that the BRCA1 locus must lie distal to the marker Mfd191 (D17S776) (Goldgar et al., 1994). This marker is known to lie distal to THRA1 and RARA. The smallest published region for the BRCA1 locus is thus between D17S776 and D17S78. This region still contains approximately 1.5 million bases of DNA, making the isolation and testing of all genes in the region a very difficult task. We have therefore undertaken the tasks of constructing a physical map of the region, isolating a set of polymorphic STR markers located in the region, and analyzing these new markers in a set of informative families to refine the location of the BRCA1 gene to a manageable interval.

Four families provide important genetic evidence for localization of BRCA1 to a sufficiently small region for the application of positional cloning strategies. Two families (K2082, K1901) provide data relating to the proximal boundary for BRCA1 and the other two (K2035, K1813) fix the distal boundary. These families are discussed in detail below. A total of 15 Short Tandem Repeat markers assayable by PCR were used to refine this localization in the families studied. These markers include DS17S7654, DS17S975, tdj1474, and tdj1239. Primer sequences for these markers are provided in SEQ ID NO:3 and SEQ ID NO:4 for

DS17S754; in SEQ ID NO:5 and SEQ ID NO:6 for DS17S975; in SEQ ID NO:7 and SEQ ID NO:8 for tdj1474; and, in SEQ ID NO:9 and SEQ ID NO:10 for tdj1239. Kindred 2082

Kindred 2082 is the largest BRCA1-linked breast/ovarian cancer family studied to date. It has a LOD score of 8.6, providing unequivocal evidence for 17q linkage. This family has been previously described and shown to contain a critical recombinant placing BRCA1 distal to MFD191 (D17S776). This recombinant occurred in a woman diagnosed with ovarian cancer at age 45 whose mother had ovarian cancer at age 63. The affected mother was deceased; however, from her children, she could be inferred to have the linked haplotype present in the 30 other linked cases in the family in the region between Mfd15 and Mfd188. Her affected daughter received the linked allele at the loci ED2, 4-7, and Mfd188, but received the allele on the non-BRCA1 chromosome at Mfd15 and Mfd91. In order to further localize this recombination breakpoint, we tested DNA from the key members of this family for the following markers derived from physical mapping resources: tdj1474, tdj1239, CF4, D17S855. For the markers tdj1474 and CF4, the affected daughter did not receive the linked allele. For the STR locus tdj1239, however, the mother could be inferred to be informative and her daughter did receive the BRCA1-associated allele. D17S855 was not informative in this family. Based on this analysis, the order is 17q centromere-Mfd191-17HSD-CF4-tdj1474-tdj1239-D17S855-ED2-4-7-Mfd188-17q telomere. The recombinant described above therefore places BRCA1 distal to tdj1474, and the breakpoint is localized to the interval between tdj1474 and tdj1239. The only alternative explanation for the data in this family other than that of BRCA1 being located distal to tdj1474, is that the ovarian cancer present in the recombinant individual is caused by reasons independent of the BRCA1 gene. Given that ovarian cancer diagnosed before age 50 is rare, this alternate explanation is exceedingly unlikely. Kindred 1901

Kindred 1901 is an early-onset breast cancer family with 7 cases of breast cancer diagnosed before 50, 4 of which were diagnosed before age 40. In addition, there were three cases of breast cancer diagnosed between the ages of 50 and 70. One case of breast cancer also had ovarian cancer at age 61. This family currently has a LOD score of 1.5 with D17S855. Given this linkage evidence and the presence of at least one ovarian cancer case, this family has a posterior probability of being due to BRCA1 of over 0.99. In this family, the recombination comes from the fact that an individual who is the brother of the ovarian cancer case from which the majority of the other cases descend, only shares a portion of the haplotype which is cosegregating with the other cases in the family. However, he passed this partial haplotype to his daughter who developed breast cancer at age 44. If this case is due to the BRCA1 gene, then only the part of the haplotype shared between this brother and his sister can contain the BRCA1 gene. The difficulty in interpretation of this kind of information is that while one can be sure of the markers which are not shared and therefore recombinant, markers which are concordant can either be shared because they are non-recombinant, or because their parent was homozygous. Without the parental genotypic data it is impossible to discriminate between these alternatives. Inspection of the haplotype in K1901, shows that he does not share the linked allele at Mfd15 (D17S250), THRA1, CF4 (D17S1320), and tdj1474 (17DS1321). He does share the linked allele at Mfd191 (D17S776), ED2 (D17S1327), tdj1239 (D17S1328), and Mfd188 (D17S579).

Although the allele shared at Mfd191 is relatively rare (0.07), we would presume that the parent was homozygous since they are recombinant with markers located nearby on either side, and a double recombination event in this region would be extremely unlikely. Thus the evidence in this family would also place the BRCA1 locus distal to tdj474. However, the lower limit of this breakpoint is impossible to determine without parental genotype information. It is intriguing that the key recombinant breakpoint in this family confirms the result in Kindred 2082. As before, the localization information in this family is only meaningful if the breast cancer was due to the BRCA1 gene. However, her relatively early age at diagnosis (44) makes this seem very likely since the risk of breast cancer before age 45 in the general population is low (approximately 1%). Kindred 2035

This family is similar to K1901 in that the information on the critical recombinant events is not directly observed but is inferred from the observation that the two haplotypes which are cosegregating with the early onset breast cancer in the two branches of the family appear identical for markers located in the proximal portion of the 17q BRCA1 region but differ at more distal loci. Each of these two haplotypes occurs in at least four cases of early-onset or bilateral breast cancer. The overall LOD score with ED2 in this family is 2.2, and considering that there is a case of ovarian cancer in the family (indicating a prior probability of BRCA1 linkage of 80%), the resulting posterior probability that this family is linked to BRCA1 is 0.998. The haplotypes are identical for the markers Mfd15, THRA1, Mfd191, ED2, AA1, D17S858 and D17S902. The common allele at Mfd15 and ED2 are both quite rare, indicating that this haplotype is shared identical by descent. The haplotypes are discordant, however, for CA375, 4-7, and Mfd188, and several more distal markers. This indicates that the BRCA1 locus must lie above the marker CA-375. This marker is located approximately 50 kb below D17S78, so it serves primarily as additional confirmation of this previous lower boundary as reported in Simard et al. (1993). Kindred 1813

Kindred 1813 is a small family with four cases of breast cancer diagnosed at very early ages whose mother also had breast cancer diagnosed at an early age and ovarian cancer some years later. This family yields a maximum multipoint LOD score of 0.60 with 17q markers and, given that there is at least one case of ovarian cancer, results in a posterior probability of being a BRCA1 linked family of 0.93. This family contains a directly observable recombination event in individual 18 (see FIG. 5 in Simard et al., *Human Mol. Genet.* 2:1193-1199 (1993)), who developed breast cancer at age 34. The genotype of her affected mother at the relevant 17q loci can be inferred from her genotypes, her affected sister's genotypes, and the genotypes of three other unaffected siblings. Individual 18 inherits the BRCA1-linked alleles for the following loci: Mfd15, THRA1, D17S800, D17S855, AA1, and D17S931. However, for markers below D17S931, i.e., U5R, vrs31, D17S858, and D17S579, she has inherited the alleles located on the non-disease bearing chromosome. The evidence from this family therefore would place the BRCA1 locus proximal to the marker U5R. Because of her early age at diagnosis (34) it is extremely unlikely that the recombinant individual's cancer is not due to the gene responsible for the other cases of breast/ovarian cancer in this family; the uncertainty in this family comes from our somewhat smaller amount of evidence that breast cancer in this family is due to BRCA1 rather than a second, as yet unmapped, breast cancer susceptibility locus.

Size of the region containing BRCA1

Based on the genetic data described in detail above, the BRCA1 locus must lie in the interval between the markers tdj1474 and U5R, both of which were isolated in our laboratory. Based upon the physical maps shown in FIGS. 2 and 3, we can try to estimate the physical distance between these two loci. It takes approximately 14 P1 clones with an average insert size of approximately 80 kb to span the region. However, because all of these P1s overlap to some unknown degree, the physical region is most likely much smaller than 14 times 80 kb. Based on restriction maps of the clones covering the region, we estimate the size of the region containing BRCA1 to be approximately 650 kb.

EXAMPLE 7

Identification of Candidate cDNA Clones for the BRCA1 Locus by Genomic Analysis of the Contig Region

Complete screen of the plausible region. The first method to identify candidate cDNAs, although labor intensive, used known techniques. The method comprised the screening of cosmids and P1 and BAC clones in the contig to identify putative coding sequences. The clones containing putative coding sequences were then used as probes on filters of cDNA libraries to identify candidate cDNA clones for future analysis. The clones were screened for putative coding sequences by either of two methods.

Zoo blots. The first method for identifying putative coding sequences was by screening the cosmid and P1 clones for sequences conserved through evolution across several species. This technique is referred to as "zoo blot analysis" and is described by Monaco, 1986. Specifically, DNAs from cow, chicken, pig, mouse and rat were digested with the restriction enzymes EcoRI and HindIII (8 µg of DNA per enzyme). The digested DNAs were separated overnight on an 0.7% gel at 20 volts for 16 hours (14 cm gel), and the DNA transferred to Nylon membranes using standard Southern blot techniques. For example, the zoo blot filter was treated at 65° C. in 0.1× SSC, 0.5% SDS, and 0.2M Tris, pH 8.0, for 30 minutes and then blocked overnight at 42° C. in 5× SSC, 10% PEG 8000, 20 mM NaPO₄ pH 6.8, 100 µg/ml Salmon Sperm DNA, 1x Denhardt's, 50% formamide, 0.1% SDS, and 2 µg/ml C₆t-1 DNA.

The cosmid and P1 clones to be analyzed were digested with a restriction enzyme to release the human DNA from the vector DNA. The DNA was separated on a 14 cm, 0.5% agarose gel run overnight at 20 volts for 16 hours. The human DNA bands were cut out of the gel and electroeluted from the gel wedge at 100 volts for at least two hours in 0.5x Tris Acetate buffer (Maniatis et al., 1982). The eluted Not I digested DNA (~15 kb to 25 kb) was then digested with EcoRI restriction enzyme to give smaller fragments (~0.5 kb to 5.0 kb) which melt apart more easily for the next step of labeling the DNA with radionucleotides. The DNA fragments were labeled by means of the hexamer random prime labeling method (Boehringer-Mannheim, Cat. #1004760). The labeled DNA was spermine precipitated (add 100 µl TE, 5 µl 0.1M spermine, and 5 µl of 10 mg/ml salmon sperm DNA) to remove unincorporated radionucleotides. The labeled DNA was then resuspended in 100 µl TE, 0.5M NaCl at 65° C. for 5 minutes and then blocked with Human C₆t-1 DNA for 2-4 hrs. as per the manufacturer's instructions (Gibco/BRL, Cat. #5279SA). The C₆t-1 blocked probe was incubated on the zoo blot filters in the blocking solution overnight at 42° C. The filters were washed for 30 minutes at room temperature in 2× SSC, 0.1% SDS, and then in the

same buffer for 30 minutes at 55° C. The filters were then exposed 1 to 3 days at -70° C. to Kodak XAR-5 film with an intensifying screen. Thus, the zoo blots were hybridized with either the pool of Eco-R1 fragments from the insert, or each of the fragments individually.

HTF island analysis. The second method for identifying cosmids to use as probes on the cDNA libraries was HTF island analysis. Since the pulsed-field map can reveal HTF islands, cosmids that map to these HTF island regions were analyzed with priority. HTF islands are segments of DNA which contain a very high frequency of unmethylated CpG dinucleotides (Tonolio et al., 1990) and are revealed by the clustering of restriction sites of enzymes whose recognition sequences include CpG dinucleotides. Enzymes known to be useful in HTF-island analysis are AscI, NotI, BssHII, EagI, SacII, NaeI, NarI, SmaI, and MluI (Anand, 1992). A pulsed-field map was created using the enzymes NotI, NruI, EagI, SacII, and SaII, and two HTF islands were found. These islands are located in the distal end of the region, one being distal to the GP2B locus, and the other being proximal to the same locus, both outside the BRCA1 region. The cosmids derived from the YACs that cover these two locations were analyzed to identify those that contain these restriction sites, and thus the HTF islands.

cDNA screening. Those clones that contain HTF islands or show hybridization to other species DNA besides human are likely to contain coding sequences. The human DNA from these clones was isolated as whole insert or as EcoRI fragments and labeled as described above. The labeled DNA was used to screen filters of various cDNA libraries under the same conditions as the zoo blots except that the cDNA filters undergo a more stringent wash of 0.1× SSC, 0.1% SDS at 65° C. for 30 minutes twice.

Most of the cDNA libraries used to date in our studies (libraries from normal breast tissue, breast tissue from a woman in her eighth month of pregnancy and a breast malignancy) were prepared at Clontech, Inc. The cDNA library generated from breast tissue of an 8 month pregnant woman is available from Clontech (Cat. #HL1037a) in the Lambda gt-10 vector, and is grown in C600Hfl bacterial host cells. Normal breast tissue and malignant breast tissue samples were isolated from a 37 year old Caucasian female and one-gram of each tissue was sent to Clontech for mRNA processing and cDNA library construction. The latter two libraries were generated using both random and oligo-dT priming, with size selection of the final products which were then cloned into the Lambda Zap II vector, and grown in XL1-blue strain of bacteria as described by the manufacturer. Additional tissue-specific cDNA libraries include human fetal brain (Stratagene, Cat. 936206), human testis (Clontech Cat. HL3024), human thymus (Clontech Cat. HL 1127n), human brain (Clontech Cat. HL11810), human placenta (Clontech Cat 1075b), and human skeletal muscle (Clontech Cat. HL1124b).

The cDNA libraries were plated with their host cells on NZCYM plates, and filter lifts are made in duplicate from each plate as per Maniatis et al. (1982). Insert (human) DNA from the candidate genomic clones was purified and radioactively labeled to high specific activity. The radioactive DNA was then hybridized to the cDNA filters to identify those cDNAs which correspond to genes located within the candidate cosmid clone. cDNAs identified by this method were picked, replated, and screened again with the labeled clone insert or its derived EcoRI fragment DNA to verify their positive status. Clones that were positive after this second round of screening were then grown up and their DNA purified for Southern blot analysis and sequencing.

5,747,282

51

Clones were either purified as plasmid through in vivo excision of the plasmid from the Lambda vector as described in the protocols from the manufacturers, or isolated from the Lambda vector as a restriction fragment and subcloned into plasmid vector.

The Southern blot analysis was performed in duplicate, one using the original genomic insert DNA as a probe to verify that cDNA insert contains hybridizing sequences. The second blot was hybridized with cDNA insert DNA from the largest cDNA clone to identify which clones represent the same gene. All cDNAs which hybridize with the genomic clone and are unique were sequenced and the DNA analyzed to determine if the sequences represent known or unique genes. All cDNA clones which appear to be unique were further analyzed as candidate BRCA1 loci. Specifically, the clones are hybridized to Northern blots to look for breast specific expression and differential expression in normal versus breast tumor RNAs. They are also analyzed by PCR on clones in the BRCA1 region to verify their location. To map the extent of the locus, full length cDNAs are isolated and their sequences used as PCR probes on the YACs and the clones surrounding and including the original identifying clones. Intron-exon boundaries are then further defined through sequence analysis.

We have screened the normal breast, 8 month pregnant breast and fetal brain cDNA libraries with zoo blot-positive Eco R1 fragments from cosmid BAC and P1 clones in the region. Potential BRCA1 cDNA clones were identified among the three libraries. Clones were picked, replated, and screened again with the original probe to verify that they were positive.

Analysis of hybrid-selected cDNA. cDNA fragments obtained from direct selection were checked by Southern blot hybridization against the probe DNA to verify that they originated from the contig. Those that passed this test were sequenced in their entirety. The set of DNA sequences obtained in this way were then checked against each other to find independent clones that overlapped. For example, the clones 694-65, 1240-1 and 1240-33 were obtained independently and subsequently shown to derive from the same contiguous cDNA sequence which has been named EST:489:1.

Analysis of candidate clones. One or more of the candidate genes generated from above were sequenced and the information used for identification and classification of each expressed gene. The DNA sequences were compared to known genes by nucleotide sequence comparisons and by translation in all frames followed by a comparison with known amino acid sequences. This was accomplished using Genetic Data Environment (GDE) version 2.2 software and the Basic Local Alignment Search Tool (Blast) series of client/server software packages (e.g., BLASTN 1.3.1 3MP), for sequence comparison against both local and remote sequence databases (e.g., GenBank), running on Sun SPARC workstations. Sequences reconstructed from collections of cDNA clones identified with the cosmids and P1s have been generated. All candidate genes that represented new sequences were analyzed further to test their candidacy for the putative BRCA1 locus.

Mutation screening. To screen for mutations in the affected pedigrees, two different approaches were followed. First, genomic DNA isolated from family members known to carry the susceptibility allele of BRCA1 was used as a template for amplification of candidate gene sequences by PCR. If the PCR primers flank or overlap an intron/exon boundary, the amplified fragment will be larger than pre-

52

dicted from the cDNA sequence or will not be present in the amplified mixture. By a combination of such amplification experiments and sequencing of P1, BAC or cosmid clones using the set of designed primers it is possible to establish the intron/exon structure and ultimately obtain the DNA sequences of genomic DNA from the pedigrees.

A second approach that is much more rapid if the intron/exon structure of the candidate gene is complex involves sequencing fragments amplified from pedigree lymphocyte cDNA. cDNA synthesized from lymphocyte mRNA extracted from pedigree blood was used as a substrate for PCR amplification using the set of designed primers. If the candidate gene is expressed to a significant extent in lymphocytes, such experiments usually produce amplified fragments that can be sequenced directly without knowledge of intron/exon junctions.

The products of such sequencing reactions were analyzed by gel electrophoresis to determine positions in the sequence that contain either mutations such as deletions or insertions, or base pair substitutions that cause amino acid changes or other detrimental effects.

Any sequence within the BRCA1 region that is expressed in breast is considered to be a candidate gene for BRCA1. Compelling evidence that a given candidate gene corresponds to BRCA1 comes from a demonstration that pedigree families contain defective alleles of the candidate.

EXAMPLE 8

Identification of BRCA1

Identification of BRCA1. Using several strategies, a detailed map of transcripts was developed for the 600 kb region of 17q21 between D17S1321 and D17S1324. Candidate expressed sequences were defined as DNA sequences obtained from: 1) direct screening of breast, fetal brain, or lymphocyte cDNA libraries, 2) hybrid selection of breast, lymphocyte or ovary cDNAs, or 3) random sequencing of genomic DNA and prediction of coding exons by XPOUND (Thomas and Skolnick, 1994). These expressed sequences in many cases were assembled into contigs composed of several independently identified sequences. Candidate genes may comprise more than one of these candidate expressed sequences. Sixty-five candidate expressed sequences within this region were identified by hybrid selection, by direct screening of cDNA libraries, and by random sequencing of P1 subclones. Expressed sequences were characterized by transcript size, DNA sequence, database comparison, expression pattern, genomic structure, and, most importantly, DNA sequence analysis in individuals from kindreds segregating 17q-linked breast and ovarian cancer susceptibility.

Three independent contigs of expressed sequence, 1141:1 (649 bp), 694:5 (213 bp) and 754:2 (1079 bp) were isolated and eventually shown to represent portions of BRCA1. When ESTs for these contigs were used as hybridization probes for Northern analysis, a single transcript of approximately 7.8 kb was observed in normal breast mRNA, suggesting that they encode different portions of a single gene. Screens of breast, fetal brain, thymus, testes, lymphocyte and placental cDNA libraries and PCR experiments with breast mRNA linked the 1141:1, 694:5 and 754:2 contigs. 5' RACE experiments with thymus, testes, and breast mRNA extended the contig to the putative 5' end, yielding a composite full length sequence. PCR and direct sequencing of P1s and BACs in the region were used to identify the location of introns and allowed the determination of splice donor and acceptor sites. These three expressed sequences were merged into a single transcription

5,747,282

53

unit that proved in the final analysis to be BRCA1. This transcription unit is located adjacent to D17S855 in the center of the 600 kb region (FIG. 4).

Combination of sequences obtained from cDNA clones, hybrid selection sequences, and amplified PCR products allowed construction of a composite full length BRCA1 cDNA (SEQ ID NO:1). The sequence of the BRCA1 cDNA (up through the stop codon) has also been deposited with GenBank and assigned accession number U-14680. This deposited sequence is incorporated herein by reference. The cDNA clone extending farthest in the 3' direction contains a poly(A) tract preceded by a polyadenylation signal. Conceptual translation of the cDNA revealed a single long open reading frame of 208 kilodaltons (amino acid sequence: SEQ ID NO:2) with a potential initiation codon flanked by sequences resembling the Kozak consensus sequence (Kozak, 1987). Smith-Waterman (Smith and Waterman, 1981) and BLAST (Altschul et al., 1990) searches identified a sequence near the amino terminus with considerable homology to zinc-finger domains (FIG. 5). This sequence contains cysteine and histidine residues present in the consensus C3HC4 zinc-finger motif and shares multiple other residues with zinc-finger proteins in the databases. The BRCA1 gene is composed of 23 coding exons arrayed over more than 100 kb of genomic DNA (FIG. 6). Northern blots using fragments of the BRCA1 cDNA as probes identified a single transcript of about 7.8 kb, present most abundantly in breast, thymus and testis, and also present in ovary (FIG. 7). Four alternatively spliced products were observed as independent cDNA clones; 3 of these were detected in breast and 2 in ovary mRNA (FIG. 6). A PCR survey from tissue cDNAs further supports the idea that there is considerable heterogeneity near the 5' end of transcripts from this gene:

54

longest deletion being 1,155 bp. The predominant form of the BRCA1 protein in breast and ovary lacks exon 4. The nucleotide sequence for BRCA1 exon 4 is shown in SEQ ID NO:11, with the predicted amino acid sequence shown in SEQ ID NO:12.

Additional 5' sequence of BRCA1 genomic DNA is set forth in SEQ ID NO:13. The G at position 1 represents the potential start site in testis. The A in position 140 represents the potential start site in somatic tissue. There are six alternative splice forms of this 5' sequence as shown in FIG. 8. The G at position 356 represents the canonical first splice donor site. The G at position 444 represents the first splice donor site in two clones (testis 1 and testis 2). The G at position 889 represents the first splice donor site in thymus 3. A fourth splice donor site is the G at position 1230. The T at position 1513 represents the splice acceptor site for all of the above splice donors. A fifth alternate splice form has a first splice donor site at position 349 with a first acceptor site at position 591 and a second splice donor site at position 889 and a second acceptor site at position 1513. A sixth alternate form is unspliced in this 5' region. The A at position 1532 is the canonical start site, which appears at position 120 of SEQ ID NO:1. Partial genomic DNA sequences determined for BRCA1 are set forth in FIGS. 10A-10H and SEQ ID Numbers:14-34. The lower case letters (in FIGS. 10A-10H) denote intron sequence while the upper case letters denote exon sequence. Indefinite intervals within introns are designated with vvvvvvvvvvvv in FIGS. 10A-10H. The intron/exon junctions are shown in Table 9. The CAG found at the 5' end of exons 8 and 14 is found in some cDNAs but not in others. Known polymorphic sites are shown in FIGS. 10A-10H in boldface type and are underlined.

TABLE 9

Exon	Base Position*			Intron Borders	
	5'	3'	Length	5'	3'
e1	1	100	100	GATAAATTAAAACTGCGACTGCGCGGGTG ^{35*}	GTAGTAGAGTCCCGGGAAGGGACAGGGGG ³⁶
e2	101	199	99	ATATATATATGTTTTCTAATGTGTAAAG ³⁷	GTAAGTCAGCACAAAGAGTGTATTAATTGG ³⁸
e3	200	253	54	TTCTTTTGTCCGCCCCCTACCGTCTAG ³⁹	GTAAGTTTGAATGTGTATGTGGCTCCATT ⁴⁰
e4	***	***	111	AGCTACTTTTTTTTTTTTTTTTGAGACAG ⁴¹	GTAAGTGCACACCACCATTATGTCAGGATAAT ⁴²
e5	254	331	78	AATGTGTTCTTTCTTTCTTATAATTATAG ⁴³	GTAATAATTGTGTAATGATGCTAGGTTGG ⁴⁴
e6	332	420	89	GAGTGTGGTTCTCAACAACATTAATTCAG ⁴⁵	GTAATGTGTAATATCCCAAGATACACAT ⁴⁶
e7	421	560	140	AAACAATAATGTTTTCCCTGTATTTACAG ⁴⁷	GTA AAAACCATTTGTTTTCTTCTTCTCTC ⁴⁸
e8	561	666	106	TGCTTGACTGTCTTTTACCATACTGTTTAA ⁴⁹	GTAAGGGTCTCAGGTTTTTTAAGTATTAA ⁵⁰
e9	667	712	46	TGATTTATTTTTTGGGGGAAATTTTATG ⁵¹	GTGAGTCAAGAGCAACCTTTTGTCATGAAG ⁵²
e10	713	789	77	TCTTATAGGACTCTGCTCTTTTCCCTATAG ⁵³	GTAATGGCAAGGTTTGCCAACTTAACAGGG ⁵⁴
e11	790	4215	3426	GAGTACCTTGTTATTTTTGTAATTTTCCAG ⁵⁵	GTAATGGCAACAGGTTTTTGTGTTTTGCCCC ⁵⁶
e12	4216	4302	87	ACATCTGAACCTCTGTTTTTGTAATTTAA ⁵⁷	AGGTAAAAGCGTGTGTGTGTGTGCACATG ⁵⁸
e13	4303	4476	174	CAITTTCTTGGTACCACTTATCGTTTTTGA ⁵⁹	GTGTGTATTTGTGGCCAAACACTGATACT ⁶⁰
e14	4477	4603	127	AGTAGATTTGTTTTCTCATTCATTAAAG ⁶¹	GTAAGAAACATCAATGTAAAGATGCTGTGG ⁶²
e15	4604	4794	191	ATGGTTTCTCCCTTCCATTATCTTCTAG ^{63**}	GTAATATTTTCACTGCTGTATTGGCAACAA ⁶⁴
e16	4795	5105	311	TGTAAATTAACACTTCTCCACTTCTTTCAG ⁶⁵	GTAGTGTATCTCATATGATCTCCCTAACT ⁶⁶
e17	5106	5193	88	ATGATAATGGAATATTTGATTTAATTTACAG ⁶⁷	GTATACCAGAAACTTTTACAGATACTCTTG ⁶⁸
e18	5194	5271	78	CTAATCCTTTGAGTGTTTTTTCATCTTCAG ⁶⁹	GTAAGTATAATACTATTTCTCCCTCCCTCC ⁷⁰
e19	5272	5312	41	GTAAACCTGTCTTTCTATGATGCTTTTAG ⁷¹	GTAAGTACTTGATGTTCAAACCTAACCCAGA ⁷²
e20	5313	5396	84	TCTGTAGTGGGTGTGTGTTGGTTTCTTCTAG ⁷³	GTAAGAGCTCCCTCCCTCAAGTTGACAAAAA ⁷⁴
e21	5397	5451	55	TCGTCCCTCTGCTCTGCTCTCTCTCCAG ⁷⁵	GTAAGAGCTGGGAGAACCCAGAGTGTCCA ⁷⁶
e22	5452	5525	74	AGTGAATTTAATCTGTAAATGTCTCAATTAG ⁷⁷	GTAAGTATTTGGGTGCGCTGTCAAGTGTGGGA ⁷⁸
e23	5526	5586	61	TTGAATGCTCTTTCCCTCTCGGGGATCCAG ⁷⁹	GTAAGGTGCGCTCAACTGTACCTGTGCTATT ⁸⁰
e24	5587	5914	328	CTAATCTTGCTTTGRTCTCTGCTCCAG ⁸¹	

*Base numbers in SEO ID NO: 1.

**Numbers in superscript refer to SEQ ID NOS.

***e4 from SEO ID NO: 11.

the molecular basis for the heterogeneity involves differential choice of the first splice donor site, and the changes detected all alter the transcript in the region 5' of the identified start codon. We have detected six potential alternate splice donors in this 5' untranslated region, with the

Low stringency blots in which genomic DNA from organisms of diverse phylogenetic background were probed with BRCA1 sequences that lack the zinc-finger region revealed strongly hybridizing fragments in human, monkey, sheep and pig, and very weak hybridization signals in rodents. This

result indicates that, apart from the zinc-finger domain, BRCA1 is conserved only at a moderate level through evolution.

Germline BRCA1 mutations in 17q-linked kindreds. The most rigorous test for BRCA1 candidate genes is to search for potentially disruptive mutations in carrier individuals from kindreds that segregate 17q-linked susceptibility to breast and ovarian cancer. Such individuals must contain BRCA1 alleles that differ from the wildtype sequence. The set of DNA samples used in this analysis consisted of DNA from individuals representing 8 different BRCA1 kindreds (Table 10).

TABLE 10

KINDRED DESCRIPTIONS AND ASSOCIATED LOD SCORES						
Kindred	Cases (n)			Sporadic Cases ¹ (n)	LOD Score	Markers(s)
	Br	< 50	Ov			
2082	31	20	22	7	9.49	D17S1327
2099	22	14	2*	0	2.36	D17S800/D17S855 ²
2035	10	8	1*	0	2.25	D17S1327
1901	10	7	1*	0	1.50	D17S855
1925	4	3	0	0	0.55	D17S579
1910	5	4	0	0	0.36	D17S579/D17S250 ²
1927	5	4	0	1	-0.44	D17S250
1911	8	5	0	2	-0.20	D17S250

¹Number of women with breast cancer (diagnosed under age 50) or ovarian cancer (diagnosed at any age) who do not share the BRCA1-linked haplotype segregating in the remainder of the cases in the kindred.
²Multipoint LOD score calculated using both markers
*kindred contains one individual who had both breast and ovarian cancer; this individual is counted as a breast cancer case and as an ovarian cancer case.

The logarithm of the odds (LOD) scores in these kindreds range from 9.49 to -0.44 for a set of markers in 17q21. Four of the families have convincing LOD scores for linkage, and 4 have low positive or negative LOD scores. The latter kindreds were included because they demonstrate haplotype sharing at chromosome 17q21 for at least 3 affected members. Furthermore, all kindreds in the set display early age of breast cancer onset and 4 of the kindreds include at least one case of ovarian cancer, both hallmarks of BRCA1 kindreds. One kindred, 2082, has nearly equal incidence of breast and ovarian cancer, an unusual occurrence given the relative rarity of ovarian cancer in the population. All of the kindreds except two were ascertained in Utah. K2035 is from the midwest. K2099 is an African-American kindred from the southern USA.

In the initial screen for predisposing mutations in BRCA1, DNA from one individual who carries the predisposition in each kindred was tested. The 23 coding exons and associated splice junction were amplified either from genomic DNA samples or from cDNA prepared from lymphocyte mRNA. When the amplified DNA sequences were compared to the wildtype sequence, 4 of the 8 kindred samples were found to contain sequence variants (Table 11).

TABLE 11

PREDISPOSING MUTATIONS			
Kindred Number	Mutation	Coding Effect	Location*
2082	C → T	Gln → Stop	4056
1910	extra C	frameshift	5385
2099	T → G	Met → Arg	5443

TABLE 11-continued

PREDISPOSING MUTATIONS			
Kindred Number	Mutation	Coding Effect	Location*
2035	?	loss of transcript	
1901	11 bp deletion	frameshift	189

*In Sequence ID NO: 1

All four sequence variants are heterozygous and each appears in only one of the kindreds. Kindred 2082 contains a nonsense mutation in exon 11 (FIG. 9A), Kindred 1910 contains a single nucleotide insertion in exon 20 (FIG. 9B), and Kindred 2099 contains a missense mutation in exon 21, resulting in a Met→Arg substitution. The frameshift and nonsense mutations are likely disruptive to the function of the BRCA1 product. The peptide encoded by the frameshift allele in Kindred 1910 would contain an altered amino acid sequence beginning 108 residues from the wildtype C-terminus. The peptide encoded by the frameshift allele in Kindred 1901 would contain an altered amino acid sequence beginning with the 24th residue from the wildtype N-terminus. The mutant allele in Kindred 2082 would encode a protein missing 551 residues from the C-terminus. The missense substitution observed in Kindred 2099 is potentially disruptive as it causes the replacement of a small hydrophobic amino acid (Met), by a large charged residue (Arg). Eleven common polymorphisms were also identified, 8 in coding sequence and 3 in introns.

The individual studied in Kindred 2035 evidently contains a regulatory mutation in BRCA 1. In her cDNA, a polymorphic site (A→G at base 3667) appeared homozygous, whereas her genomic DNA revealed heterozygosity at this position (FIG. 9C). A possible explanation for this observation is that mRNA from her mutated BRCA1 allele is absent due to a mutation that affects its production or stability. This possibility was explored further by examining 5 polymorphic sites in the BRCA1 coding region, which are separated by as much as 3.5 kb in the BRCA1 transcript. In all cases where her genomic DNA appeared heterozygous for a polymorphism, cDNA appeared homozygous. In individuals from other kindreds and in non-haplotype carriers in Kindred 2035, these polymorphic sites could be observed as heterozygous in cDNA, implying that amplification from cDNA was not biased in favor of one allele. This analysis indicates that a BRCA1 mutation in Kindred 2035 either prevents transcription or causes instability or aberrant splicing of the BRCA1 transcript.

Cosegregation of BRCA1 mutations with BRCA1 haplotypes and population frequency analysis. In addition to potentially disrupting protein function, two criteria must be met for a sequence variant to qualify as a candidate predisposing mutation. The variant must: 1) be present in individuals from the kindred who carry the predisposing BRCA 1 haplotype and absent in other members of the kindred, and 2) be rare in the general population.

Each mutation was tested for cosegregation with BRCA1. For the frameshift mutation in Kindred 1910, two other haplotype carriers and one non-carrier were sequenced (FIG. 9B). Only the carriers exhibited the frameshift mutation. The C to T change in Kindred 2082 created a new AvrII restriction site. Other carriers and non-carriers in the kindred were tested for the presence of the restriction site (FIG. 9A). An allele-specific oligonucleotide (ASO) was designed to detect the presence of the sequence variant in Kindred 2099. Several individuals from the kindred, some known to carry

the haplotype associated with the predisposing allele, and others known not to carry the associated haplotype, were screened by ASO for the mutation previously detected in the kindred. In each kindred, the corresponding mutant allele was detected in individuals carrying the BRCA1-associated haplotype, and was not detected in noncarriers. In the case of the potential regulatory mutation observed in the individual from Kindred 2035, cDNA and genomic DNA from carriers in the kindred were compared for heterozygosity at polymorphic sites. In every instance, the extinguished allele in the cDNA sample was shown to lie on the chromosome that carries the BRCA1 predisposing allele (FIG. 9C).

To exclude the possibility that the mutations were simply common polymorphisms in the population, ASOs for each mutation were used to screen a set of normal DNA samples. Gene frequency estimates in Caucasians were based on random samples from the Utah population. Gene frequency estimates in African-Americans were based on 39 samples provided by M. Peracek-Vance which originate from African-Americans used in her linkage studies and 20 newborn Utah African-Americans. None of the 4 potential predisposing mutations was found in the appropriate control population, indicating that they are rare in the general population. Thus, two important requirements for BRCA1 susceptibility alleles were fulfilled by the candidate predisposing mutations: 1) cosegregation of the mutant allele with disease, and 2) absence of the mutant allele in controls, indicating a low gene frequency in the general population.

Phenotypic Expression of BRCA1 Mutations. The effect of the mutations on the BRCA1 protein correlated with differences in the observed phenotypic expression in the BRCA1 kindreds. Most BRCA1 kindreds have a moderately increased ovarian cancer risk, and a smaller subset have high risks of ovarian cancer, comparable to those for breast cancer (Easton et al., 1993). Three of the four kindreds in which BRCA1 mutations were detected fall into the former category, while the fourth (K2082) falls into the high ovarian cancer risk group. Since the BRCA1 nonsense mutation found in K2082 lies closer to the amino terminus than the other mutations detected, it might be expected to have a different phenotype. In fact, Kindred K2082 mutation has a high incidence of ovarian cancer, and a later mean age at diagnosis of breast cancer cases than the other kindreds (Goldgar et al., 1994). This difference in age of onset could be due to an ascertainment bias in the smaller, more highly penetrant families, or it could reflect tissue-specific differences in the behavior of BRCA1 mutations. The other 3 kindreds that segregate known BRCA1 mutations have, on average, one ovarian cancer for every 10 cases of breast cancer, but have a high proportion of breast cancer cases diagnosed in their late 20's or early 30's. Kindred 1910, which has a frameshift mutation, is noteworthy because three of the four affected individuals had bilateral breast cancer, and in each case the second tumor was diagnosed within a year of the first occurrence. Kindred 2035, which segregates a potential regulatory BRCA1 mutation, might also be expected to have a dramatic phenotype. Eighty percent of breast cancer cases in this kindred occur under age 50. This figure is as high as any in the set, suggesting a BRCA1 mutant allele of high penetrance (Table 10).

Although the mutations described above clearly are deleterious, causing breast cancer in women at very young ages, each of the four kindreds with mutations includes at least one woman who carries the mutation who lived until age 80 without developing a malignancy. It will be of utmost importance in the studies that follow to identify other genetic or environmental factors that may ameliorate the effects of BRCA1 mutations.

In four of the eight putative BRCA1-linked kindreds, potential predisposing mutations were not found. Three of these four have LOD scores for BRCA1-linked markers of less than 0.55. Thus, these kindreds may not in reality segregate BRCA1 predisposing alleles. Alternatively, the mutations in these four kindreds may lie in regions of BRCA1 that, for example, affect the level of transcript and therefore have thus far escaped detection.

Role of BRCA1 in Cancer. Most tumor suppressor genes identified to date give rise to protein products that are absent, nonfunctional, or reduced in function. The majority of TP53 mutations are missense; some of these have been shown to produce abnormal p53 molecules that interfere with the function of the wildtype product (Shaulian et al., 1992; Srivastava et al., 1993). A similar dominant negative mechanism of action has been proposed for some adenomatous polyposis coli (APC) alleles that produce truncated molecules (Su et al., 1993), and for point mutations in the Wilms' tumor gene (WT1) that alter DNA binding of the protein (Little et al., 1993). The nature of the mutations observed in the BRCA1 coding sequence is consistent with production of either dominant negative proteins or nonfunctional proteins. The regulatory mutation inferred in Kindred 2035 cannot be a dominant negative; rather, this mutation likely causes reduction or complete loss of BRCA1 expression from the affected allele.

The BRCA1 protein contains a C₃HC₄ zinc-finger domain, similar to those found in numerous DNA binding proteins and implicated in zinc-dependent binding to nucleic acids. The first 180 amino acids of BRCA1 contain five more basic residues than acidic residues. In contrast, the remainder of the molecule is very acidic, with a net excess of 70 acidic residues. The excess negative charge is particularly concentrated near the C-terminus. Thus, one possibility is that BRCA1 encodes a transcription factor with an N-terminal DNA binding domain and a C-terminal transactivational "acidic blob" domain. Interestingly, another familial tumor suppressor gene, WT1, also contains a zinc-finger motif (Haber et al., 1990). Many cancer predisposing mutations in WT1 alter zinc-finger domains (Little et al., 1993; Haber et al., 1990; Little et al., 1992). WT1 encodes a transcription factor, and alternative splicing of exons that encode parts of the zinc-finger domain alter the DNA binding properties of WT1 (Bickmore et al., 1992). Some alternatively spliced forms of WT1 mRNA generate molecules that act as transcriptional repressors (Drummond et al., 1994). Some BRCA1 splicing variants may alter the zinc-finger motif, raising the possibility that a regulatory mechanism similar to 10 that which occurs in WT1 may apply to BRCA1.

EXAMPLE 9

Analysis of Tumors for BRCA1 Mutations

To focus the analysis on tumors most likely to contain BRCA1 mutations, primary breast and ovarian carcinomas were typed for LOH in the BRCA1 region. Three highly polymorphic, simple tandem repeat markers were used to assess LOH: D17S1323 and D17S855, which are intragenic to BRCA1, and D17S1327, which lies approximately 100 kb distal to BRCA1. The combined LOH frequency in informative cases (i.e., where the germline was heterozygous) was 32/72 (44%) for the breast carcinomas and 12/21 (57%) for the ovarian carcinomas, consistent with previous measurements of LOH in the region (Futreal et al., 1992b; Jacobs et al., 1993; Sato et al., 1990; Eccles et al., 1990; Cropp et al., 1994). The analysis thus defined a panel of 32

breast tumors and 12 ovarian tumors of mixed race and age of onset to be examined for BRCA mutations. The complete 5,589 bp coding region and intron/exon boundary sequences of the gene were screened in this tumor set by direct sequencing alone or by a combination of single-strand conformation analysis (SSCA) and direct sequencing.

A total of six mutations (of which two are identical) was found, one in an ovarian tumor, four in breast tumors and one in a male unaffected haplotype carrier (Table 12). One mutation, Glu 1541Ter, introduced a stop codon that would create a truncated protein missing 323 amino acids at the carboxy terminus. In addition, two missense mutations were identified. These are Ala1708Glu and Met1775Arg and involve substitutions of small, hydrophobic residues by charged residues. Patients 17764 and 19964 are from the same family. In patient OV24 nucleotide 2575 is deleted and in patients 17764 and 19964 nucleotides 2993–2996 are deleted.

TABLE 12

Predisposing Mutations					
Patient	Codon	Nucleotide Change	Amino Acid Change	Age of Onset	Family History
BT098	1541	GAG → TAG	Glu → Stop	39	–
OV24	819	1 bp deletion	frameshift	44	–
BT106	1708	GCG → GAG	Ala → Glu	24	+
MC44	1775	ATG → AGG	Met → Arg	42	+
17764	958	4 bp deletion	frameshift	31	+
19964	958	4 bp deletion	frameshift		++

*Unaffected haplotype carrier, male

Several lines of evidence suggest that all five mutations represent BRCA1 susceptibility alleles:

- (i) all mutations are present in the germline;
- (ii) all are absent in appropriate control populations, suggesting they are not common polymorphisms;
- (iii) each mutant allele is retained in the tumor, as is the case in tumors from patients belonging to kindreds that segregate BRCA1 susceptibility alleles (Smith et al., 1992; Kelsell et al., 1993) (if the mutations represented neutral polymorphisms, they should be retained in only 50% of the cases);
- (iv) the age of onset in the four breast cancer cases with mutations varied between 24 and 42 years of age, consistent with the early age of onset of breast cancer in individuals with BRCA1 susceptibility; similarly, the ovarian cancer case was diagnosed at 44, an age that falls in the youngest 13% of all ovarian cancer cases; and finally,
- (v) three of the five cases have positive family histories of breast or ovarian cancer found retrospectively in their medical records, although the tumor set was not selected with regard to this criterion.

BT106 was diagnosed at a very early age with breast cancer. Her mother had ovarian cancer, her father had melanoma, and her paternal grandmother also had breast cancer. Patient MC44, an African-American, had bilateral breast cancer at an early age. This patient had a sister who died of breast cancer at a very early age. Her mutation (Met1775Arg) had been detected previously in Kindred 2099, an African-American family that segregates a BRCA1 susceptibility allele, and was absent in African-American and Caucasian controls.

Patient MC44, to our knowledge, is unrelated to Kindred 2099. The detection of a rare mutant allele, once in a BRCA1

kindred and once in the germline of an apparently unrelated early-onset breast cancer case, suggests that the Met1775Arg change may be a common predisposing mutation in African-Americans. Collectively, these observations indicate that all four BRCA1 mutations in tumors represent susceptibility alleles; no somatic mutations were detected in the samples analyzed.

The paucity of somatic BRCA1 mutations is unexpected, given the frequency of LOH on 17q, and the usual role of susceptibility genes as tumor suppressors in cancer progression. There are three possible explanations for this result: (i) some BRCA1 mutations in coding sequences were missed by our screening procedure; (ii) BRCA1 somatic mutations fall primarily outside the coding exons; and (iii) LOH events in 17q do not reflect BRCA1 somatic mutations.

If somatic BRCA1 mutations truly are rare in breast and ovary carcinomas, this would have strong implications for the biology of BRCA1. The apparent lack of somatic BRCA1 mutations implies that there may be some fundamental difference in the genesis of tumors in genetically predisposed BRCA1 carriers, compared with tumors in the general population. For example, mutations in BRCA1 may have an effect only on tumor foination at a specific stage early in breast and ovarian development. This possibility is consistent with a primary function for BRCA1 in premenopausal breast cancer. Such a model for the role of BRCA1 in breast and ovarian cancer predicts an interaction between reproductive hormones and BRCA1 function. However, no clinical or pathological differences in familial versus sporadic breast and ovary tumors, other than age of onset, have been described (Lynch et al., 1990). On the other hand, the recent finding of increased TP53 mutation and microsatellite instability in breast tumors from patients with a family history of breast cancer (Glebov et al., 1994) may reflect some difference in tumors that arise in genetically predisposed persons. The involvement of BRCA1 in this phenomenon can now be addressed directly. Alternatively, the lack of somatic BRCA1 mutations may result from the existence of multiple genes that function in the same pathway of tumor suppression as BRCA1, but which collectively represent a more favored target for mutation in sporadic tumors. Since mutation of a single element in a genetic pathway is generally sufficient to disrupt the pathway, BRCA1 might mutate at a rate that is far lower than the sum of the mutational rates of the other elements.

EXAMPLE 10

Analysis of the BRCA1 Gene

The structure and function of BRCA1 gene are determined according to the following methods.

Biological Studies. Mammalian expression vectors containing BRCA1 cDNA are constructed and transfected into appropriate breast carcinoma cells with lesions in the gene. Wild-type BRCA1 cDNA as well as altered BRCA1 cDNA are utilized. The altered BRCA1 cDNA can be obtained from altered BRCA1 alleles or produced as described below. Phenotypic reversion in cultures (e.g., cell morphology, doubling time, anchorage-independent growth) and in animals (e.g., tumorigenicity) is examined. The studies will employ both wild-type and mutant forms (Section B) of the gene.

Molecular Genetics Studies. In vitro mutagenesis is performed to construct deletion mutants and missense mutants (by single base-pair substitutions in individual codons and cluster charged→alanine scanning mutagenesis). The mutants are used in biological, biochemical and biophysical studies.

5,747,282

61

Mechanism Studies. The ability of BRCA1 protein to bind to known and unknown DNA sequences is examined. Its ability to transactivate promoters is analyzed by transient reporter expression systems in mammalian cells. Conventional procedures such as particle-capture and yeast two-hybrid system are used to discover and identify any functional partners. The nature and functions of the partners are characterized. These partners in turn are targets for drug discovery.

Structural Studies. Recombinant proteins are produced in *E. coli*, yeast, insect and/or mammalian cells and are used in crystallographical and NMR studies. Molecular modeling of the proteins is also employed. These studies facilitate structure-driven drug design.

EXAMPLE 11

Two Step Assay to Detect the Presence of BRCA1 in a Sample

Patient sample is processed according to the method disclosed by Antonarakis et al. (1985), separated through a 1% agarose gel and transferred to nylon membrane for Southern blot analysis.

Membranes are UV cross linked at 150 mJ using a GS Gene Linker (Bio-Rad). BRCA1 probe corresponding to nucleotide positions 3631–3930 of SEQ ID NO:1 is subcloned into pTZ18U. The phagemids are transformed into *E. coli* MV1190 infected with M13KO7 helper phage (Bio-Rad, Richmond, Calif.). Single stranded DNA is isolated according to standard procedures (see Sambrook et al., 1989).

Blots are prehybridized for 15–30 min at 65° C. in 7% sodium dodecyl sulfate (SDS) in 0.5M NaPO₄. The methods follow those described by Nguyen et al., 1992. The blots are hybridized overnight at 65° C. in 7% SDS, 0.5M NaPO₄ with 25–50 ng/ml single stranded probe DNA. Post-hybridization washes consist of two 30 min washes in 5% SDS, 40 mM NaPO₄ at 65° C., followed by two 30 min washes in 1% SDS, 40 mM NaPO₄ at 65° C.

Next the blots are rinsed with phosphate buffered saline (pH 6.8) for 5 min at room temperature and incubated with 0.2% casein in PBS for 30–60 min at room temperature and rinsed in PBS for 5 min. The blots are then preincubated for 5–10 minutes in a shaking water bath at 45° C. with hybridization buffer consisting of 6M urea, 0.3M NaCl, and 5X Denhardt's solution (see Sambrook, et al., 1989). The buffer is removed and replaced with 50–75 µl/cm² fresh hybridization buffer plus 2.5 nM of the covalently cross-linked oligonucleotide-alkaline phosphatase conjugate with the nucleotide sequence complementary to the universal primer site (UP-AP, Bio-Rad). The blots are hybridized for 20–30 min at 45° C. and post hybridization washes are incubated at 45° C. as two 10 min washes in 6M urea, 1× standard saline citrate (SSC), 0.1% SDS and one 10 min wash in 1× SSC, 0.1% Triton®X-100. The blots are rinsed for 10 min at room temperature with 1× SSC.

Blots are incubated for 10 min at room temperature with shaking in the substrate buffer consisting of 0.1M diethanolamine, 1 mM MgCl₂, 0.02% sodium azide, pH 10.0. Individual blots are placed in heat sealable bags with substrate buffer and 0.2 mM AMPPD (3-(2'-spiroadamantane)-4-methoxy-4-(3'-phosphoryloxy)phenyl-1,2-dioxetane, disodium salt, Bio-Rad). After a 20 min incubation at room temperature with shaking, the excess AMPPD solution is removed. The blot is exposed to X-ray film overnight. Positive bands indicate the presence of BRCA1.

62

EXAMPLE 12

Generation of Polyclonal Antibody against BRCA1

Segments of BRCA1 coding sequence were expressed as fusion protein in *E. coli*. The overexpressed protein was purified by gel elution and used to immunize rabbits and mice using a procedure similar to the one described by Harlow and Lane, 1988. This procedure has been shown to generate Abs against various other proteins (for example, see Kraemer et al., 1993).

Briefly, a stretch of BRCA1 coding sequence was cloned as a fusion protein in plasmid PET5A (Novagen, Inc., Madison, Wis.). The BRCA1 incorporated sequence includes the amino acids corresponding to #1361–1554 of SEQ ID NO:2. After induction with IPTG, the overexpression of a fusion protein with the expected molecular weight was verified by SDS/PAGE. Fusion protein was purified from the gel by electroelution. The identification of the protein as the BRCA1 fusion product was verified by protein sequencing at the N-terminus. Next, the purified protein was used as immunogen in rabbits. Rabbits were immunized with 100 µg of the protein in complete Freund's adjuvant and boosted twice in 3 week intervals, first with 100µg of immunogen in incomplete Freund's adjuvant followed by 100 µg of immunogen in PBS. Antibody containing serum is collected two weeks thereafter.

This procedure is repeated to generate antibodies against the mutant forms of the BRCA1 gene.

These antibodies, in conjunction with antibodies to wild type BRCA1, are used to detect the presence and the relative level of the mutant forms in various tissues and biological fluids. EXAMPLE 13

Generation of Monoclonal Antibodies Specific for BRCA1

Monoclonal antibodies are generated according to the following protocol. Mice are immunized with immunogen comprising intact BRCA1 or BRCA1 peptides (wild type or mutant) conjugated to keyhole limpet hemocyanin using glutaraldehyde or EDC as is well known.

The immunogen is mixed with an adjuvant. Each mouse receives four injections of 10 to 100 µg of immunogen and after the fourth injection blood samples are taken from the mice to determine if the serum contains antibody to the immunogen. Serum titer is determined by ELISA or RIA. Mice with sera indicating the presence of antibody to the immunogen are selected for hybridoma production.

Spleens are removed from immune mice and a single cell suspension is prepared (see Harlow and Lane, 1988). Cell fusions are performed essentially as described by Kohler and Milstein, 1975. Briefly, P3.65.3 myeloma cells (American Type Culture Collection, Rockville, Md.) are fused with immune spleen cells using polyethylene glycol as described by Harlow and Lane, 1988. Cells are plated at a density of 2×10⁵ cells/well in 96 well tissue culture plates. Individual wells are examined for growth and the supernatants of wells with growth are tested for the presence of BRCA1 specific antibodies by ELISA or RIA using wild type or mutant BRCA1 target protein. Cells in positive wells are expanded and subcloned to establish and confirm monoclonality.

Clones with the desired specificities are expanded and grown as ascites in mice or in a hollow fiber system to produce sufficient quantities of antibody for characterization and assay development.

A000195

EXAMPLE 14

Sandwich Assay for BRCA1

It will be appreciated that the methods and compositions of the instant invention can be incorporated in the form of a

64

LIST OF REFERENCES

66

- A000197**

5,747,282

67

Sheffield, V. C., et al. (1991). *Am. J. Hum. Genet.* 49:699-706.
Shenk, et al. (1975). *Proc. Natl. Acad. Sci. USA* 72:989.
Shimada, et al. (1991). *J. Clin. Invest.* 88:1043-1047.
Shinkai, Y., et al. (1992). *Cell* 68:855.
Shizuya, H., et al. (1992). *Proc. Natl. Acad. Sci. USA* 89:8794-8797.
Simard, J., et al. (1993). *Human Mol Genet.* 2:1193-1199.
Skolnick, M. H. and Wallace, B. R. (1988). *Genomics* 2:273-279.
Skolnick, M. H., et al. (1990). *Science* 250:1715-1720.
Smith, S. A., et al. (1992). *Nature Genetics* 2:128-131.
Smith, T. F. and Waterman, M.S. (1981). *J. Mol. Biol.* 147:195-197.
Snouwaert, J. N., et al. (1992). *Science* 257:1083.
Sorge, et al. (1984). *Mol. Cell. Biol.* 4:1730-1737.
Srivastava, S., et al. (1993). *Cancer Res.* 53:4452-5.
Sternberg (1990). *Proc. Natl. Acad. Sci. USA* 87:103-107.
Stemberg, et al. (1990). *The New Biologist* 2:151-162.
Stewart, et al. (1992). *Hum. Gene Ther.* 3:267-275.
Stratford-Perricaudet, et al. (1990). *Hum. Gene Ther.* 1:241-256.
Swift, M., et al. (1991). *N. Engl. J. Med.* 325:1831-1836.
Swift, M., et al. (1976). *Cancer Res.* 36:209-215.
Su, L. K., et al. (1993). *Cancer Res.* 53:2728-31.
Thomas, A. and Skolnick, M. H. (1994). *IMA Journal of Mathematics Applied in Medicine and Biology* (in press).
Tonolio, D., et al. (1990). Cold Spring Harbor Conference.
Valancius, V. & Smithies, O. (1991). *Mol. Cell Biol.* 11:1402.
van Dilla, et al. (1986). *Biotechnology* 4:537-552.
Wagner, et al. (1990). *Proc. Natl. Acad. Sci. USA* 87:3410-3414.
Wagner, et al. (1991). *Proc. Natl. Acad. Sci. USA* 88:4255-4259.
Wang and Huang (1989). *Biochemistry* 28:9508-9514.
Wartell, R. M., et al. (1990). *Nucl. Acids Res.* 18:2699-2705.
Weber, J. L. (1990). *Genomics* 7:524-530.

68

Weber and May (1989). *Am. J. Hum. Genet.* 44:388-396.
Weber, J. L., et al. (1990). *Nucleic Acid Res.* 18:4640.
Wells, J. A. (1991). *Methods in Enzymol.* 202:390-411.
Wetmur & Davidson (1968). *J. Mol. Biol.* 31:349-370.
5 White, M. B., et al. (1992). *Genomics* 12:301-306.
White and LaUuel (1988). *Ann. Rev. Genet.* 22:259-279.
Wilkinson, et al. (1992). *Nucleic Acids Res.* 20:224-2239.
Williams and Anderson (1984). *Genet. Epidemiol.* 1:7-20.
Wolff, et al. (1990). *Science* 247:1465-1468.
10 Wolff, et al. (1991). *BioTechniques* 11:474-485.
Wooster, R., et al. (1994). *Science* 265:2088.
Wu, et al. (989a). *Genomics* 4:560-569.
Wu, et al. (989b). *J. Biol. Chem.* 264:16985-16987.
15 Wu, et al. (991). *J. Biol. Chem.* 266:143 138-14342.
Zenke, et al. (1 990). *Proc. Natl. Acad. Sci. USA* 87:3655-3659.
List of Patents and Patent Applications :
U.S. Pat. No. 3,817,837
20 U.S. Pat. No. 3,850,752
U.S. Pat. No. 3,939,350
U.S. Pat. No. 3,996,345
U.S. Pat. No. 4,275,149
U.S. Pat. No. 4,277,437
25 U.S. Pat. No. 4,366,241
U.S. Pat. No. 4,376,110
U.S. Pat. No. 4,486,530
U.S. Pat. No. 4,683,195
U.S. Pat. No. 4,683,202
30 U.S. Pat. No. 4,816,567
U.S. Pat. No. 4,868,105
U.S. Pat. No. 5,252,479
EPO Publication No. 225,807
European Patent Application Publication No. 0332435
35 Geysen, H., PCT published application WO 84/03564, published 13 Sep. 1984
Hitzeman et al., EP 73,675A
PCT published application WO 93/07282

SEQUENCE LISTING

(1) GENERAL INFORMATION:
(i i i) NUMBER OF SEQUENCES: 85
(2) INFORMATION FOR SEQ ID NO:1:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5914 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear
(i i) MOLECULE TYPE: cDNA
(i i i) HYPOTHETICAL: NO
(i v) ANTI-SENSE: NO
(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens
(i x) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 120..5708
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:1:

AGCTCGCTGA GACTTCCTGG ACCCCGCACC AGGCTGTGGG GTTTCACAGA TAACTGGGCC 60

5,747,282

69

70

-continued

CCTGCGCTCA GGAGGCCTTC ACCCTCTGCT CTGGGTAAAG TTCATTGGAA CAGAAAGAA																119
ATG Met 1	GAT Asp	TTA Leu	TCT Ser	GCT Ala 5	CTT Leu	CGC Arg	GTT Val	GAA Glu	GAA Glu 10	GTA Val	CAA Gln	AAT Asn	GTC Val	ATT Ile 15	AAT Asn	167
GCT Ala	ATG Met	CAG Gln	AAA Lys 20	ATC Ile	TTA Leu	GAG Glu	TGT Cys	CCC Pro 25	ATC Ile	TGT Cys	CTG Leu	GAG Glu	TTG Leu 30	ATC Ile	AAG Lys	215
GAA Glu	CCT Pro	GTC Val 35	TCC Ser	ACA Thr	AAG Lys	TGT Cys	GAC Asp 40	CAC His	ATA Ile	TTT Phe	TGC Cys	AAA Lys 45	TTT Phe	TGC Cys	ATG Met	263
CTG Leu	AAA Lys 50	CTT Leu	CTC Leu	AAC Asn	CAG Gln	AAG Lys 55	AAA Lys	GGG Gly	CCT Pro	TCA Ser	CAG Gln 60	TGT Cys	CCT Pro	TTA Leu	TGT Cys	311
AAG Lys 65	AAT Asn	GAT Asp	ATA Ile	ACC Thr	AAA Lys 70	AGG Arg	AGC Ser	CTA Leu	CAA Gln	GAA Glu 75	AGT Ser	ACG Thr	AGA Arg	TTT Phe	AGT Ser 80	359
CAA Gln	CTT Leu	GTT Val	GAA Glu 85	GAG Glu	CTA Leu	TTG Leu	AAA Lys	ATC Ile 90	ATT Ile	TGT Cys	GCT Ala	TTT Phe	CAG Gln	CTT Leu 95	GAC Asp	407
ACA Thr	GGT Gly	TTG Leu	GAG Glu 100	TAT Tyr	GCA Ala	AAC Asn	AGC Ser	TAT Tyr 105	AAT Asn	TTT Phe	GCA Ala	AAA Lys	AAG Lys 110	GAA Glu	AAT Asn	455
AAC Asn	TCT Ser	CCT Pro 115	GAA Glu	CAT His	CTA Leu	AAA Lys	GAT Asp 120	GAA Glu	GTT Val	TCT Ser	ATC Ile 125	ATC Ile	CAA Gln	AGT Ser	ATG Met	503
GGC Gly	TAC Tyr	AGA Arg	AAC Asn	CGT Arg	GCC Ala	AAA Lys 135	AGA Arg	CTT Leu	CTA Leu	CAG Gln	AGT Ser 140	GAA Glu	CCC Pro	GAA Glu	AAT Asn	551
CCT Pro 145	TCC Ser	TTG Leu	CAG Gln	GAA Glu 150	ACC Thr	AGT Ser	CTC Leu	AGT Ser	GTC Val	CAA Gln 155	CTC Leu	TCT Ser	AAC Asn	CTT Leu 160	GGA Gly	599
ACT Thr	GTG Val	AGA Arg	ACT Thr	CTG Leu 165	AGG Arg	ACA Thr	AAG Lys	CAG Gln	CGG Arg 170	ATA Ile	CAA Gln	CCT Pro	CAA Gln	AAG Lys 175	ACG Thr	647
TCT Ser	GTC Val	TAC Tyr	ATT Ile 180	GAA Glu	TTG Leu	GGA Gly	TCT Ser	GAT Asp 185	TCT Ser	TCT Ser	GAA Glu	GAT Asp	ACC Thr 190	GTT Val	AAT Asn	695
AAG Lys	GCA Ala	ACT Thr 195	TAT Tyr	TGC Cys	AGT Ser	GTG Val	GGA Gly 200	GAT Asp	CAA Gln	GAA Glu	TTG Leu 205	TTA Leu	CAA Gln	ATC Ile	ACC Thr	743
CCT Pro	CAA Gln 210	GGA Gly	ACC Thr	AGG Arg	GAT Asp	GAA Glu 215	ATC Ile	AGT Ser	TTG Leu	GAT Asp	TCT Ser 220	GCA Ala	AAA Lys	AAG Lys	GCT Ala	791
GCT Ala 225	TGT Cys	GAA Glu	TTT Phe	TCT Ser	GAG Glu 230	ACG Thr	GAT Asp	GTA Val	ACA Thr	AAT Asn 235	ACT Thr	GAA Glu	CAT His	CAT His	CAA Gln 240	839
CCC Pro	AGT Ser	AAT Asn	AAT Asn 245	GAT Asp	TTG Asn	AAC Asn	ACC Thr	ACT Thr	GAG Glu 250	AAG Lys	CGT Arg	GCA Ala	GCT Ala	GAG Glu 255	AGG Arg	887
CAT His	CCA Pro	GAA Glu	AAG Lys 260	TAT Tyr	CAG Gln	GGT Gly	AGT Ser	TCT Ser 265	GTT Val	TCA Ser	AAC Asn	TTG Leu	CAT His 270	GTG Val	GAG Glu	935
CCA Pro	TGT Cys 275	GGC Gly	ACA Thr	AAT Asn	ACT Thr	CAT His	GCC Ala 280	AGC Ser	TCA Ser	TTA Leu	CAG Gln	CAT His 285	GAG Glu	AAC Asn	AGC Ser	983
AGT Ser	TTA Leu 290	TTA Leu	CTC Leu	ACT Thr	AAA Lys	GAC Asp 295	AGA Arg	ATG Met	AAT Asn	GTA Val	GAA Glu 300	AAG Lys	GCT Ala	GAA Glu	TTC Phe	1031
TGT	AAT	AAA	AGC	AAA	CAG	CCT	GGC	TTA	GCA	AGG	AGC	CAA	CAT	AAC	AGA	1077

A000199

5,747,282

71

72

-continued

Cys 305	Asn	Lys	Ser	Lys	Gln 310	Pro	Gly	Leu	Ala	Arg 315	Ser	Gln	His	Asn	Arg 320	
TGG	GCT	GGA	AGT	AAG	GAA	ACA	TGT	AAT	GAT	AGG	CGG	ACT	CCC	AGC	ACA	1127
Trp	Ala	Gly	Ser	Lys 325	Glu	Thr	Cys	Asn	Asp 330	Arg	Arg	Thr	Pro	Ser 335	Thr	
GAA	AAA	AAG	GTA	GAT	CTG	AAT	GCT	GAT	CCC	CTG	TGT	GAG	AGA	AAA	GAA	1175
Glu	Lys	Lys	Val	Asp	Leu	Asn	Ala	Asp 345	Pro	Leu	Cys	Glu	Arg 350	Lys	Glu	
TGG	AAT	AAG	CAG	AAA	CTG	CCA	TGC	TCA	GAG	AAT	CCT	AGA	GAT	ACT	GAA	1223
Trp	Asn	Lys 355	Gln	Lys	Leu	Pro	Cys	Ser 360	Glu	Asn	Pro	Arg 365	Asp	Thr	Glu	
GAT	GTT	CCT	TGG	ATA	ACA	CTA	AAT	AGC	AGC	ATT	CAG	AAA	GTT	AAT	GAG	1271
Asp	Val 370	Pro	Trp	Ile	Thr	Leu 375	Asn	Ser	Ser	Ile	Gln 380	Lys	Val	Asn	Glu	
TGG	TTT	TCC	AGA	AGT	GAT	GAA	CTG	TTA	GGT	TCT	GAT	GAC	TCA	CAT	GAT	1319
Trp	Phe	Ser	Arg	Ser	Asp 390	Glu	Leu	Leu	Gly	Ser 395	Asp	Asp	Ser	His	Asp 400	
GGG	GAG	TCT	GAA	TCA	AAT	GCC	AAA	GTA	GCT	GAT	GTA	TTG	GAC	GTT	CTA	1367
Gly	Glu	Ser	Glu	Ser 405	Asn	Ala	Lys	Val	Ala 410	Asp	Val	Leu	Asp 415	Val	Leu	
AAT	GAG	GTA	GAT	GAA	TAT	TCT	GGT	TCT	TCA	GAG	AAA	ATA	GAC	TTA	CTG	1415
Asn	Glu	Val	Asp	Glu	Tyr	Ser	Gly	Ser 425	Ser	Glu	Lys	Ile	Asp 430	Leu	Leu	
GCC	AGT	GAT	CCT	CAT	GAG	GCT	TTA	ATA	TGT	AAA	AGT	GAA	AGA	GTT	CAC	1463
Ala	Ser	Asp 435	Pro	His	Glu	Ala	Leu 440	Ile	Cys	Lys	Ser	Glu 445	Arg	Val	His	
TCC	AAA	TCA	GTA	GAG	AGT	AAT	ATT	GAA	GAC	AAA	ATA	TTT	GGG	AAA	ACC	1511
Ser	Lys 450	Ser	Val	Glu	Ser	Asn 455	Ile	Glu	Asp	Lys	Ile 460	Phe	Gly	Lys	Thr	
TAT	CGG	AAG	AAG	GCA	AGC	CTC	CCC	AAC	TTA	AGC	CAT	GTA	ACT	GAA	AAT	1559
Tyr	Arg	Lys	Lys	Ala	Ser 470	Leu	Pro	Asn	Leu	Ser 475	His	Val	Thr	Glu	Asn 480	
CTA	ATT	ATA	GGA	GCA	TTT	GTT	ACT	GAG	CCA	CAG	ATA	ATA	CAA	GAG	CGT	1607
Leu	Ile	Ile	Gly	Ala 485	Phe	Val	Thr	Glu	Pro 490	Gln	Ile	Ile	Gln	Glu 495	Arg	
CCC	CTC	ACA	AAT	AAA	TTA	AAG	CGT	AAA	AGG	AGA	CCT	ACA	TCA	GGC	CTT	1655
Pro	Leu	Thr	Asn 500	Lys	Leu	Lys	Arg	Lys 505	Arg	Arg	Pro	Thr	Ser 510	Gly	Leu	
CAT	CCT	GAG	GAT	TTT	ATC	AAG	AAA	GCA	GAT	TTG	GCA	GTT	CAA	AAG	ACT	1703
His	Pro	Glu 515	Asp	Phe	Ile	Lys	Lys 520	Ala	Asp	Leu	Ala 525	Val	Gln	Lys	Thr	
CCT	GAA	ATG	ATA	AAT	CAG	GGA	ACT	AAC	CAA	ACG	GAG	CAG	AAT	GGT	CAA	1751
Pro	Glu 530	Met	Ile	Asn	Gln	Gly 535	Thr	Asn	Gln	Thr	Glu 540	Gln	Asn	Gly	Gln	
GTG	ATG	AAT	ATT	ACT	AAT	AGT	GGT	CAT	GAG	AAT	AAA	ACA	AAA	GGT	GAT	1799
Val	Met	Asn	Ile	Thr	Asn 550	Ser	Gly	His	Glu	Asn 555	Lys	Thr	Lys	Gly	Asp 560	
TCT	ATT	CAG	AAT	GAG	AAA	AAT	CCT	AAC	CCA	ATA	GAA	TCA	CTC	GAA	AAA	1847
Ser	Ile	Gln	Asn 565	Lys	Lys	Asn	Pro	Asn	Pro 570	Ile	Glu	Ser	Leu	Glu 575	Lys	
GAA	TCT	GCT	TTC	AAA	ACG	AAA	GCT	GAA	CCT	ATA	AGC	AGC	AGT	ATA	AGC	1895
Glu	Ser	Ala	Phe	Lys	Thr	Lys	Ala	Glu 585	Pro	Ile	Ser	Ser	Ser 590	Ile	Ser	
AAT	ATG	GAA	CTC	GAA	TTA	AAT	ATC	CAC	AAT	TCA	AAA	GCA	CCT	AAA	AAG	1943
Asn	Met	Glu 595	Leu	Glu	Leu	Asn 600	Ile	His	Asn	Ser	Lys	Ala 605	Pro	Lys	Lys	
AAT	AGG	CTG	AGG	AGG	AAG	TCT	TCT	ACC	AGG	CAT	ATT	CAT	GCG	CTT	GAA	1991
Asn	Arg 610	Leu	Arg	Arg	Lys	Ser 615	Ser	Thr	Arg	His	Ile 620	His	Ala	Leu	Glu	
CTA	GTA	GTC	AGT	AGA	AAT	CTA	AGC	CCA	CCT	AAT	TGT	ACT	GAA	TTG	CAA	2039

5,747,282

73

74

-continued

Leu 625	Val	Val	Ser	Arg	Asn 630	Leu	Ser	Pro	Pro	Asn 635	Cys	Thr	Glu	Leu	Gln 640	
ATT Ile	GAT Asp	AGT Ser	TGT Cys	TCT Ser 645	AGC Ser	AGT Ser	GAA Glu	GAG Glu	ATA Ile 650	AAG Lys	AAA Lys	AAA Lys	AAG Lys	TAC Tyr 655	AAC Asn	2087
CAA Gln	ATG Met	CCA Pro	GTC Val 660	AGG Arg	CAC His	AGC Ser	AGA Arg	AAC Asn 665	CTA Leu	CAA Gln	CTC Leu	ATG Met	GAA Glu 670	GGT Gly	AAA Lys	2135
GAA Glu	CCT Pro	GCA Ala 675	ACT Thr	GGA Gly	GCC Ala	AAG Lys	AAG Lys 680	AGT Ser	AAC Asn	AAG Lys	CCA Pro	AAT Asn 685	GAA Glu	CAG Gln	ACA Thr	2183
AGT Ser	AAA Lys 690	AGA Arg	CAT His	GAC Asp	AGC Ser	GAT Asp 695	ACT Thr	TTC Phe	CCA Pro	GAG Glu	CTG Leu 700	AAG Lys	TTA Leu	ACA Thr	AAT Asn	2231
GCA Ala 705	CCT Pro	GGT Gly	TCT Ser	TTT Phe	ACT Thr 710	AAG Lys	TGT Cys	TCA Ser	AAT Asn	ACC Thr 715	AGT Ser	GAA Glu	CTT Leu	AAA Lys	GAA Glu 720	2279
TTT Phe	GTC Val	AAT Asn	CCT Pro	AGC Ser 725	CTT Leu	CCA Pro	AGA Arg	GAA Glu 730	GAA Glu	AAA Lys	GAA Glu	GAG Glu	AAA Lys	CTA Leu 735	GAA Glu	2327
ACA Thr	GTT Val	AAA Lys	GTG Val 740	TCT Ser	AAT Asn	AAT Asn	GCT Ala	GAA Glu 745	GAC Asp	CCC Pro	AAA Lys	GAT Asp	CTC Leu 750	ATG Met	TTA Leu	2375
AGT Ser	GGA Gly	GAA Glu 755	AGG Arg	GTT Val	TTG Leu	CAA Gln	ACT Thr 760	GAA Glu	AGA Arg	TCT Ser	GTA Val	GAG Glu 765	AGT Ser	AGC Ser	AGT Ser	2423
ATT Ile	TCA Ser	TTG Leu	GTA Val	CCT Pro	GGT Gly	ACT Thr 775	GAT Asp	TAT Tyr	GGC Gly	ACT Thr	CAG Gln 780	GAA Glu	AGT Ser	ATC Ile	TCG Ser	2471
TTA Leu 785	CTG Leu	GAA Glu	GTT Val	AGC Ser	ACT Thr 790	CTA Leu	GGG Gly	AAG Lys	GCA Ala 795	AAA Lys	ACA Thr	GAA Glu	CCA Pro	AAT Asn	AAA Lys 800	2519
TGT Cys	GTG Val	AGT Ser	CAG Gln	TGT Cys 805	GCA Ala	GCA Ala	TTT Phe	GAA Glu	AAC Asn 810	CCC Pro	AAG Lys	GGA Gly	CTA Leu	ATT Ile 815	CAT His	2567
GGT Gly	TGT Cys	TCC Ser	AAA Lys 820	GAT Asp	AAT Asn	AGA Arg	AAT Asn	GAC Asp 825	ACA Thr	GAA Glu	GGC Gly	TTT Phe	AAG Lys 830	TAT Tyr	CCA Pro	2615
TTG Leu	GGA Gly	CAT His 835	GAA Glu	GTT Val	AAC Asn	CAC His	AGT Ser 840	CGG Arg	GAA Glu	ACA Thr	AGC Ser	ATA Ile 845	GAA Glu	ATG Met	GAA Glu	2663
GAA Glu	AGT Ser	GAA Glu	CTT Leu	GAT Asp	GCT Ala	CAG Gln	TAT Tyr	TTG Leu	CAG Gln	AAT Asn	ACA Thr 860	TTC Phe	AAG Lys	GTT Val	TCA Ser	2711
AAG Lys 865	CGC Arg	CAG Gln	TCA Ser	TTT Phe	GCT Ala 870	CCG Pro	TTT Phe	TCA Ser	AAT Asn	CCA Pro 875	GGA Gly	AAT Asn	GCA Ala	GAA Glu	GAG Glu 880	2759
GAA Glu	TGT Cys	GCA Ala	ACA Thr	TTC Phe 885	TCT Ser	GCC Ala	CAC His	TCT Ser	GGG Gly 890	TCC Ser	TTA Leu	AAG Lys	AAA Lys	CAA Gln 895	AGT Ser	2807
CCA Pro	AAA Lys	GTG Val	ACT Thr 900	TTT Phe	GAA Glu	TGT Cys	GAA Glu	CAA Gln 905	AAG Lys	GAA Glu	GAA Glu	AAT Asn	CAA Gln 910	GGA Gly	AAG Lys	2855
AAT Asn	GAG Glu	TCT Ser 915	AAT Asn	ATC Ile	AAG Lys	CCT Pro	GTA Val 920	CAG Gln	ACA Thr	GTT Val	AAT Asn	ATC Ile 925	ACT Thr	GCA Ala	GGC Gly	2903
TTT Phe	CCT Pro	GTG Val	GTT Val	GGT Gly	CAG Gln	AAA Lys 935	GAT Asp	AAG Lys	CCA Pro	GTT Val	GAT Asp 940	AAT Asn	GCC Ala	AAA Lys	TGT Cys	2951
AGT	ATC	AAA	GGA	GGC	TCT	AGG	TTT	TGT	CTA	TCA	TCT	CAG	TTC	AGA	GGC	2999

A000201

5,747,282

75

76

-continued

Ser 945	Ile	Lys	Gly	Gly	Ser 950	Arg	Phe	Cys	Leu	Ser 955	Ser	Gln	Phe	Arg	Gly 960	
AAC Asn	GAA Glu	ACT Thr	GGA Gly	CTC Leu 965	ATT Ile	ACT Thr	CCA Pro	AAT Asn	AAA Lys 970	CAT His	GGA Gly	CTT Leu	TTA Leu	CAA Gln 975	AAC Asn	3047
CCA Pro	TAT Tyr	CGT Arg	ATA Ile 980	CCA Pro	CCA Pro	CTT Leu	TTT Phe	CCC Pro 985	ATC Ile	AAG Lys	TCA Ser	TTT Phe	GTT Val 990	AAA Lys	ACT Thr	3095
AAA Lys	TGT Cys	AAG Lys 995	AAA Lys	AAT Asn	CTG Leu	CTA Leu	GAG Glu 1000	GAA Glu	AAC Asn	TTT Phe	GAG Glu	GAA Glu 1005	CAT His	TCA Ser	ATG Met	3143
TCA Ser	CCT Pro	GAA Glu 1010	AGA Arg	GAA Glu	ATG Met	GGA Gly 1015	AAT Asn	GAG Glu	AAC Asn	ATT Ile	CCA Pro 1020	AGT Ser	ACA Thr	GTG Val	AGC Ser	3191
ACA Thr 1025	ATT Ile	AGC Ser	CGT Arg	AAT Asn	AAC Asn 1030	ATT Ile	AGA Arg	GAA Glu	AAT Asn	GTT Val 1035	TTT Phe	AAA Lys	GAA Glu	GCC Ala	AGC Ser 1040	3239
TCA Ser	AGC Ser	AAT Asn	ATT Ile	AAT Asn 1045	GAA Glu	GTA Val	GGT Gly	TCC Ser	AGT Ser 1050	ACT Thr	AAT Asn	GAA Glu	GTG Val	GGC Gly 1055	TCC Ser	3287
AGT Ser	ATT Ile	AAT Asn 1060	GAA Glu 1060	ATA Ile	GGT Gly	TCC Ser	AGT Ser	GAT Asp 1065	GAA Glu	AAC Asn	ATT Ile	CAA Gln 1070	GCA Ala	GAA Glu	CTA Leu	3335
GGT Gly	AGA Arg	AAC Asn 1075	AGA Arg	GGG Gly	CCA Pro	AAA Lys	TTG Leu 1080	AAT Asn	GCT Ala	ATG Met	CTT Leu	AGA Arg 1085	TTA Leu	GGG Gly	GTT Val	3383
TTG Leu	CAA Gln 1090	CCT Pro	GAG Glu	GTC Val	TAT Tyr	AAA Lys 1095	CAA Gln	AGT Ser	CTT Leu	CCT Pro	GGA Gly 1100	AGT Ser	AAT Asn	TGT Cys	AAG Lys	3431
CAT His 1105	CCT Pro	GAA Glu	ATA Ile	AAA Lys	AAG Lys 1110	CAA Gln	GAA Glu	TAT Tyr	GAA Glu	GAA Glu 1115	GTA Val	GTT Val	CAG Gln	ACT Thr	GTT Val 1120	3479
AAT Asn	ACA Thr	GAT Asp	TTC Phe 1125	TCT Ser	CCA Pro	TAT Tyr	CTG Leu	ATT Ile	TCA Ser 1130	GAT Asp	AAC Asn	TTA Leu	GAA Glu	CAG Gln 1135	CCT Pro	3527
ATG Met	GGA Gly	AGT Ser	AGT Ser 1140	CAT His	GCA Ala	TCT Ser	CAG Gln	GTT Val 1145	TGT Cys	TCT Ser	GAG Glu	ACA Thr	CCT Pro 1150	GAT Asp	GAC Asp	3575
CTG Leu	TTA Leu	GAT Asp 1155	GAT Asp	GGT Gly	GAA Glu	ATA Ile	AAG Lys 1160	GAA Glu	GAT Asp	ACT Thr	AGT Ser	TTT Phe 1165	GCT Ala	GAA Glu	AAT Asn	3623
GAC Asp	ATT Ile 1170	AAG Lys	GAA Glu	AGT Ser	TCT Ser 1175	GCT Ala	GTT Val 1175	TTT Phe	AGC Ser	AAA Lys 1180	AGC Ser 1180	GTC Val	CAG Gln	AAA Lys	GGA Gly	3671
GAG Glu 1185	CTT Leu	AGC Ser	AGG Arg	AGT Ser	CCT Pro 1190	AGC Ser	CCT Pro	TTC Phe	ACC Thr	CAT His 1195	ACA Thr	CAT His	TTG Leu	GCT Ala	CAG Gln 1200	3719
GGT Gly	TAC Tyr	CGA Arg	AGA Arg	GGG Gly 1205	GCC Ala	AAG Lys	AAA Lys	TTA Leu 1210	GAG Glu 1210	TCC Ser	TCA Ser	GAA Glu	GAG Glu 1215	AAC Asn	TTA Leu	3767
TCT Ser	AGT Ser	GAG Glu	GAT Asp 1220	GAA Glu	GAG Glu	CTT Leu	CCC Pro	TGC Cys 1225	TTC Phe	CAA Gln	CAC His	TTG Leu	TTA Leu 1230	TTT Phe	GGT Gly	3815
AAA Lys	GTA Val 1235	AAC Asn	AAT Asn	ATA Ile	CCT Pro	TCT Ser	CAG Gln 1240	TCT Ser	ACT Thr	AGG Arg	CAT His	AGC Ser 1245	ACC Thr	GTT Val	GCT Ala	3863
ACC Thr	GAG Glu 1250	TGT Cys	CTG Leu	TCT Ser	AAG Lys 1255	AAC Asn	ACA Thr	GAG Glu	GAG Glu	AAT Asn 1260	TTA Leu	TTA Leu	TCA Ser	TTG Leu	AAG Lys	3911
AAT	AGC	TTA	AAT	GAC	TGC	AGT	AAC	CAG	GTA	ATA	TTG	GCA	AAG	GCA	TCT	3959

A000202

5,747,282

77

78

-continued

Asn 1265	Ser 1265	Leu 1265	Asn 1265	Asp 1265	Cys 1270	Ser 1270	Asn 1270	Gln 1270	Val 1270	Ile 1275	Leu 1275	Ala 1275	Lys 1275	Ala 1280	Ser 1280
CAG	GAA	CAT	CAC	CTT	AGT	GAG	GAA	ACA	AAA	TGT	TCT	GCT	AGC	TTG	TTT
Gln	Glu	His	His	Leu	Ser	Glu	Glu	Thr	Lys	Cys	Ser	Ala	Ser	Leu	Phe
				1285					1290					1295	
TCT	TCA	CAG	TGC	AGT	GAA	TTG	GAA	GAC	TTG	ACT	GCA	AAT	ACA	AAC	ACC
Ser	Ser	Gln	Cys	Ser	Glu	Leu	Glu	Asp	Leu	Thr	Ala	Asn	Thr	Asn	Thr
			1300					1305					1310		
CAG	GAT	CCT	TTC	TTG	ATT	GGT	TCT	TCC	AAA	CAA	ATG	AGG	CAT	CAG	TCT
Gln	Asp	Pro	Phe	Leu	Ile	Gly	Ser	Ser	Lys	Gln	Met	Arg	His	Gln	Ser
		1315					1320					1325			
GAA	AGC	CAG	GGA	GTT	GGT	CTG	AGT	GAC	AAG	GAA	TTG	GTT	TCA	GAT	GAT
Glu	Ser	Gln	Gly	Val	Gly	Leu	Ser	Asp	Lys	Glu	Leu	Val	Ser	Asp	Asp
	1330					1335					1340				
GAA	GAA	AGA	GGA	ACG	GGC	TTG	GAA	GAA	AAT	AAT	CAA	GAA	GAG	CAA	AGC
Glu	Glu	Arg	Gly	Thr	Gly	Leu	Glu	Glu	Asn	Asn	Gln	Glu	Glu	Gln	Ser
	1345				1350					1355					1360
ATG	GAT	TCA	AAC	TTA	GGT	GAA	GCA	GCA	TCT	GGG	TGT	GAG	AGT	GAA	ACA
Met	Asp	Ser	Asn	Leu	Gly	Glu	Ala	Ala	Ser	Gly	Cys	Glu	Ser	Glu	Thr
				1365					1370					1375	
AGC	GTC	TCT	GAA	GAC	TGC	TCA	GGG	CTA	TCC	TCT	CAG	AGT	GAC	ATT	TTA
Ser	Val	Ser	Gln	Asp	Cys	Ser	Gly	Leu	Ser	Ser	Gln	Ser	Asp	Ile	Leu
			1380					1385					1390		
ACC	ACT	CAG	CAG	AGG	GAT	ACC	ATG	CAA	CAT	AAC	CTG	ATA	AAG	CTC	CAG
Thr	Thr	Gln	Gln	Arg	Asp	Thr	Met	Gln	His	Asn	Leu	Ile	Lys	Leu	Gln
		1395					1400					1405			
CAG	GAA	ATG	GCT	GAA	CTA	GAA	GCT	GTG	TTA	GAA	CAG	CAT	GGG	AGC	CAG
Gln	Glu	Met	Ala	Glu	Leu	Glu	Ala	Val	Leu	Glu	Gln	His	Gly	Ser	Gln
	1410					1415					1420				
CCT	TCT	AAC	AGC	TAC	CCT	TCC	ATC	ATA	AGT	GAC	TCT	TCT	GCC	CTT	GAG
Pro	Ser	Asn	Ser	Tyr	Pro	Ser	Ile	Ile	Ser	Asp	Ser	Ser	Ala	Leu	Glu
	1425				1430					1435					1440
GAC	CTG	CGA	AAT	CCA	GAA	CAA	AGC	ACA	TCA	GAA	AAA	GCA	GTA	TTA	ACT
Asp	Leu	Arg	Asn	Pro	Glu	Gln	Ser	Thr	Ser	Glu	Lys	Ala	Val	Leu	Thr
				1445					1450					1455	
TCA	CAG	AAA	AGT	AGT	GAA	TAC	CCT	ATA	AGC	CAG	AAT	CCA	GAA	GGC	CTT
Ser	Gln	Lys	Ser	Ser	Glu	Tyr	Pro	Ile	Ser	Gln	Asn	Pro	Glu	Gly	Leu
			1460					1465					1470		
TCT	GCT	GAC	AAG	TTT	GAG	GTG	TCT	GCA	GAT	AGT	TCT	ACC	AGT	AAA	AAT
Ser	Ala	Asp	Lys	Phe	Glu	Val	Ser	Ala	Asp	Ser	Ser	Thr	Ser	Lys	Asn
		1475					1480					1485			
AAA	GAA	CCA	GGA	GTG	GAA	AGG	TCA	TCC	CCT	TCT	AAA	TGC	CCA	TCA	TTA
Lys	Glu	Pro	Gly	Val	Glu	Arg	Ser	Ser	Pro	Ser	Lys	Cys	Pro	Ser	Leu
	1490					1495					1500				
GAT	GAT	AGG	TGG	TAC	ATG	CAC	AGT	TGC	TCT	GGG	AGT	CTT	CAG	AAT	AGA
Asp	Asp	Arg	Trp	Tyr	Met	His	Ser	Cys	Ser	Gly	Ser	Leu	Gln	Asn	Arg
	1505				1510					1515					1520
AAC	TAC	CCA	TCT	CAA	GAG	GAG	CTC	ATT	AAG	GTT	GTT	GAT	GTG	GAG	GAG
Asn	Tyr	Pro	Ser	Gln	Glu	Glu	Leu	Ile	Lys	Val	Val	Asp	Val	Glu	Glu
				1525					1530					1535	
CAA	CAG	CTG	GAA	GAG	TCT	GGG	CCA	CAC	GAT	TTG	ACG	GAA	ACA	TCT	TAC
Gln	Gln	Leu	Glu	Glu	Ser	Gly	Pro	His	Asp	Leu	Thr	Glu	Thr	Ser	Tyr
			1540					1545					1550		
TTG	CCA	AGG	CAA	GAT	CTA	GAG	GGA	ACC	CCT	TAC	CTG	GAA	TCT	GGA	ATC
Leu	Pro	Arg	Gln	Asp	Leu	Glu	Gly	Thr	Pro	Tyr	Leu	Glu	Ser	Gly	Ile
		1555					1560					1565			
AGC	CTC	TTC	TCT	GAT	GAC	CCT	GAA	TCT	GAT	CCT	TCT	GAA	GAC	AGA	GCC
Ser	Leu	Phe	Ser	Asp	Asp	Pro	Glu	Ser	Asp	Pro	Ser	Glu	Asp	Arg	Ala
	1570					1575					1580				
CCA	GAG	TCA	GCT	CGT	GTT	GGC	AAC	ATA	CCA	TCT	TCA	ACC	TCT	GCA	TTG

A000203

5,747,282

79

80

-continued

Pro	Glu	Ser	Ala	Arg	Val	Gly	Asn	Ile	Pro	Ser	Thr	Ser	Ala	Leu		
1585					1590				1595					1600		
AAA	GTT	CCC	CAA	TTG	AAA	GTT	GCA	GAA	TCT	GCC	CAG	AGT	CCA	GCT	GCT	4967
Lys	Val	Pro	Gln	Leu	Lys	Val	Ala	Glu	Ser	Ala	Gln	Ser	Pro	Ala	Ala	
				1605					1610					1615		
GCT	CAT	ACT	ACT	GAT	ACT	GCT	GGG	TAT	AAT	GCA	ATG	GAA	GAA	AGT	GTG	5015
Ala	His	Thr	Thr	Asp	Thr	Ala	Gly	Tyr	Asn	Ala	Met	Glu	Glu	Ser	Val	
				1620				1625					1630			
AGC	AGG	GAG	AAG	CCA	GAA	TTG	ACA	GCT	TCA	ACA	GAA	AGG	GTC	AAC	AAA	5063
Ser	Arg	Glu	Lys	Pro	Glu	Leu	Thr	Ala	Ser	Thr	Glu	Arg	Val	Asn	Lys	
		1635					1640					1645				
AGA	ATG	TCC	ATG	GTG	GTG	TCT	GGC	CTG	ACC	CCA	GAA	GAA	TTT	ATG	CTC	5111
Arg	Met	Ser	Met	Val	Val	Ser	Gly	Leu	Thr	Pro	Glu	Glu	Phe	Met	Leu	
	1650					1655					1660					
GTG	TAC	AAG	TTT	GCC	AGA	AAA	CAC	CAC	ATC	ACT	TTA	ACT	AAT	CTA	ATT	5159
Val	Tyr	Lys	Phe	Ala	Arg	Lys	His	His	Ile	Thr	Leu	Thr	Asn	Leu	Ile	
	1665				1670				1675					1680		
ACT	GAA	GAG	ACT	ACT	CAT	GTT	GTT	ATG	AAA	ACA	GAT	GCT	GAG	TTT	GTG	5207
Thr	Glu	Glu	Thr	Thr	His	Val	Val	Met	Lys	Thr	Asp	Ala	Glu	Phe	Val	
				1685					1690					1695		
TGT	GAA	CGG	ACA	CTG	AAA	TAT	TTT	CTA	GGA	ATT	GCG	GGA	GGA	AAA	TGG	5255
Cys	Glu	Arg	Thr	Leu	Lys	Tyr	Phe	Leu	Gly	Ile	Ala	Gly	Gly	Lys	Trp	
			1700					1705					1710			
GTA	GTT	AGC	TAT	TTC	TGG	GTG	ACC	CAG	TCT	ATT	AAA	GAA	AGA	AAA	ATG	5303
Val	Val	Ser	Tyr	Phe	Trp	Val	Thr	Gln	Ser	Ile	Lys	Glu	Arg	Lys	Met	
		1715					1720					1725				
CTG	AAT	GAG	CAT	GAT	TTT	GAA	GTC	AGA	GGA	GAT	GTG	GTC	AAT	GGA	AGA	5351
Leu	Asn	Glu	His	Asp	Phe	Glu	Val	Arg	Gly	Asp	Val	Val	Asn	Gly	Arg	
	1730				1735						1740					
AAC	CAC	CAA	GGT	CCA	AAG	CGA	GCA	AGA	GAA	TCC	CAG	GAC	AGA	AAG	ATC	5399
Asn	His	Gln	Gly	Pro	Lys	Arg	Ala	Arg	Glu	Ser	Gln	Asp	Arg	Lys	Ile	
	1745				1750					1755					1760	
TTC	AGG	GGG	CTA	GAA	ATC	TGT	TGC	TAT	GGG	CCC	TTC	ACC	AAC	ATG	CCC	5447
Phe	Arg	Gly	Leu	Glu	Ile	Cys	Cys	Tyr	Gly	Pro	Phe	Thr	Asn	Met	Pro	
				1765					1770					1775		
ACA	GAT	CAA	CTG	GAA	TGG	ATG	GTA	CAG	CTG	TGT	GGT	GCT	TCT	GTG	GTG	5495
Thr	Asp	Gln	Leu	Glu	Trp	Met	Val	Gln	Leu	Cys	Gly	Ala	Ser	Val	Val	
			1780					1785					1790			
AAG	GAG	CTT	TCA	TCA	TTC	ACC	CTT	GGC	ACA	GGT	GTC	CAC	CCA	ATT	GTG	5543
Lys	Glu	Leu	Ser	Ser	Phe	Thr	Thr	Gly	Thr	Gly	Val	His	Pro	Ile	Val	
		1795					1800					1805				
GTT	GTG	CAG	CCA	GAT	GCC	TGG	ACA	GAG	GAC	AAT	GGC	TTC	CAT	GCA	ATT	5591
Val	Val	Gln	Pro	Asp	Ala	Trp	Thr	Glu	Asp	Asn	Gly	Phe	His	Ala	Ile	
		1810				1815					1820					
GGG	CAG	ATG	TGT	GAG	GCA	CCT	GTG	GTG	ACC	GCA	GAG	TGG	GTG	TTG	GAC	5639
Gly	Gln	Met	Cys	Glu	Ala	Pro	Val	Val	Thr	Arg	Glu	Trp	Val	Leu	Asp	
		1825			1830					1835				1840		
AGT	GTA	GCA	CTC	TAC	CAG	TGC	CAG	GAG	CTG	GAC	ACC	TAC	CTG	ATA	CCC	5687
Ser	Val	Ala	Leu	Tyr	Gln	Cys	Gln	Glu	Leu	Asp	Thr	Tyr	Leu	Ile	Pro	
				1845					1850					1855		
CAG	ATC	CCC	CAC	AGC	CAC	TAC	TGA	CTGCAGCCAG	CCACAGGTAC	AGAGCCACAG						5741
Gln	Ile	Pro	His	Ser	His	Tyr										
			1860													
GACCCCAAGA	ATGAGCTTAC	AAAGTGGCCT	TTCCAGGCCC	TGGGAGCTCC	TCTCACTCTT											5801
CAGTCCTTCT	ACTGTCCTGG	CTACTAAATA	TTTTATGTAC	ATCAGCCTGA	AAAGGACTTC											5861
TGGCTATGCA	AGGGTCCCTT	AAAGATTITC	TGCTTGAAGT	CTCCCTTGGG	AAT											5914

(2) INFORMATION FOR SEO ID NO:2:

A000204

5,747,282

81

82

-continued

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1863 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(i i) MOLECULE TYPE: protein

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met 1	Asp	Leu	Ser	Ala 5	Leu	Arg	Val	Glu	Glu 10	Val	Gln	Asn	Val	Ile 15	Asn
Ala	Met	Gln	Lys 20	Ile	Leu	Glu	Cys	Pro 25	Ile	Cys	Leu	Glu	Leu 30	Ile	Lys
Glu	Pro	Val 35	Ser	Thr	Lys	Cys	Asp 40	His	Ile	Phe	Cys	Lys 45	Phe	Cys	Met
Leu	Lys 50	Leu	Leu	Asn	Gln	Lys 55	Lys	Gly	Pro	Ser	Gln 60	Cys	Pro	Leu	Cys
Lys 65	Asn	Asp	Ile	Thr	Lys 70	Arg	Ser	Leu	Gln	Glu 75	Ser	Thr	Arg	Phe	Ser 80
Gln	Leu	Val	Glu	Glu 85	Leu	Leu	Lys	Ile	Ile 90	Cys	Ala	Phe	Gln	Leu 95	Asp
Thr	Gly	Leu	Glu 100	Tyr	Ala	Asn	Ser	Tyr 105	Asn	Phe	Ala	Lys	Lys 110	Glu	Asn
Asn	Ser	Pro 115	Glu	His	Leu	Lys	Asp 120	Glu	Val	Ser	Ile	Ile 125	Gln	Ser	Met
Gly	Tyr 130	Arg	Asn	Arg	Ala	Lys 135	Arg	Leu	Leu	Gln	Ser 140	Glu	Pro	Glu	Asn
Pro 145	Ser	Leu	Gln	Glu	Thr 150	Ser	Leu	Ser	Val	Gln 155	Leu	Ser	Asn	Leu	Gly 160
Thr	Val	Arg	Thr	Leu 165	Arg	Thr	Lys	Gln	Arg 170	Ile	Gln	Pro	Gln	Lys 175	Thr
Ser	Val	Tyr	Ile 180	Glu	Leu	Gly	Ser	Asp 185	Ser	Ser	Glu	Asp	Thr 190	Val	Asn
Lys	Ala	Thr 195	Tyr	Cys	Ser	Val	Gly 200	Asp	Gln	Glu	Leu	Leu 205	Gln	Ile	Thr
Pro 210	Gln	Gly	Thr	Arg	Asp 215	Glu	Ile	Ser	Leu	Asp	Ser 220	Ala	Lys	Lys	Ala
Ala 225	Cys	Glu	Phe	Ser	Glu 230	Thr	Asp	Val	Thr	Asn 235	Thr	Glu	His	His	Gln 240
Pro	Ser	Asn	Asn	Asp 245	Leu	Asn	Thr	Thr	Glu 250	Lys	Arg	Ala	Ala	Glu 255	Arg
His	Pro	Glu	Lys 260	Tyr	Gln	Gly	Ser	Ser 265	Val	Ser	Asn	Leu	His 270	Val	Glu
Pro	Cys	Gly 275	Thr	Asn	Thr	His	Ala 280	Ser	Ser	Leu	Gln	His 285	Glu	Asn	Ser
Ser 290	Leu	Leu	Leu	Thr	Lys	Asp 295	Arg	Met	Asn	Val	Glu 300	Lys	Ala	Glu	Phe
Cys 305	Asn	Lys	Ser	Lys	Gln 310	Pro	Gly	Leu	Ala	Arg 315	Ser	Gln	His	Asn	Arg 320
Trp	Ala	Gly	Ser	Lys 325	Glu	Thr	Cys	Asn	Asp 330	Arg	Arg	Thr	Pro	Ser 335	Thr
Glu	Lys	Lys	Val 340	Asp	Leu	Asn	Ala	Asp 345	Pro	Leu	Cys	Glu	Arg 350	Lys	Glu
Trp	Asn	Lys 355	Gln	Lys	Leu	Pro	Cys 360	Ser	Glu	Asn	Pro	Arg 365	Asp	Thr	Glu
Asp 370	Val	Pro	Trp	Ile	Thr	Leu 375	Asn	Ser	Ser	Ile	Gln 380	Lys	Val	Asn	Glu

A000205

5,747,282

83

84

-continued

Trp	Phe	Ser	Arg	Ser	Asp	Glu	Leu	Leu	Gly	Ser	Asp	Asp	Ser	His	Asp
385					390					395					400
Gly	Glu	Ser	Glu	Ser	Asn	Ala	Lys	Val	Ala	Asp	Val	Leu	Asp	Val	Leu
				405					410					415	
Asn	Glu	Val	Asp	Glu	Tyr	Ser	Gly	Ser	Ser	Glu	Lys	Ile	Asp	Leu	Leu
			420					425					430		
Ala	Ser	Asp	Pro	His	Glu	Ala	Leu	Ile	Cys	Lys	Ser	Glu	Arg	Val	His
		435					440					445			
Ser	Lys	Ser	Val	Glu	Ser	Asn	Ile	Glu	Asp	Lys	Ile	Phe	Gly	Lys	Thr
	450					455					460				
Tyr	Arg	Lys	Lys	Ala	Ser	Leu	Pro	Asn	Leu	Ser	His	Val	Thr	Glu	Asn
465					470					475					480
Leu	Ile	Ile	Gly	Ala	Phe	Val	Thr	Glu	Pro	Gln	Ile	Ile	Gln	Glu	Arg
			485					490					495		
Pro	Leu	Thr	Asn	Lys	Leu	Lys	Arg	Lys	Arg	Arg	Pro	Thr	Ser	Gly	Leu
			500					505					510		
His	Pro	Glu	Asp	Phe	Ile	Lys	Lys	Ala	Asp	Leu	Ala	Val	Gln	Lys	Thr
		515					520					525			
Pro	Glu	Met	Ile	Asn	Gln	Gly	Thr	Asn	Gln	Thr	Glu	Gln	Asn	Gly	Gln
	530					535					540				
Val	Met	Asn	Ile	Thr	Asn	Ser	Gly	His	Glu	Asn	Lys	Thr	Lys	Gly	Asp
545					550					555				560	
Ser	Ile	Gln	Asn	Glu	Lys	Asn	Pro	Asn	Pro	Ile	Glu	Ser	Leu	Glu	Lys
			565					570					575		
Glu	Ser	Ala	Phe	Lys	Thr	Lys	Ala	Glu	Pro	Ile	Ser	Ser	Ser	Ile	Ser
		580						585					590		
Asn	Met	Glu	Leu	Glu	Leu	Asn	Ile	His	Asn	Ser	Lys	Ala	Pro	Lys	Lys
		595				600						605			
Asn	Arg	Leu	Arg	Arg	Lys	Ser	Ser	Thr	Arg	His	Ile	His	Ala	Leu	Glu
	610					615					620				
Leu	Val	Val	Ser	Arg	Asn	Leu	Ser	Pro	Pro	Asn	Cys	Thr	Glu	Leu	Gln
625					630					635					640
Ile	Asp	Ser	Cys	Ser	Ser	Ser	Glu	Glu	Ile	Lys	Lys	Lys	Lys	Tyr	Asn
			645					650						655	
Gln	Met	Pro	Val	Arg	His	Ser	Arg	Asn	Leu	Gln	Leu	Met	Glu	Gly	Lys
			660					665					670		
Glu	Pro	Ala	Thr	Gly	Ala	Lys	Lys	Ser	Asn	Lys	Pro	Asn	Glu	Gln	Thr
		675					680					685			
Ser	Lys	Arg	His	Asp	Ser	Asp	Thr	Phe	Pro	Glu	Leu	Lys	Leu	Thr	Asn
	690					695					700				
Ala	Pro	Gly	Ser	Phe	Thr	Lys	Cys	Ser	Asn	Thr	Ser	Glu	Leu	Lys	Glu
705					710					715					720
Phe	Val	Asn	Pro	Ser	Leu	Pro	Arg	Glu	Glu	Lys	Glu	Glu	Lys	Leu	Glu
			725					730						735	
Thr	Val	Lys	Val	Ser	Asn	Asn	Ala	Glu	Asp	Pro	Lys	Asp	Leu	Met	Leu
			740					745					750		
Ser	Gly	Glu	Arg	Val	Leu	Gln	Thr	Glu	Arg	Ser	Val	Glu	Ser	Ser	Ser
		755					760					765			
Ile	Ser	Leu	Val	Pro	Gly	Thr	Asp	Tyr	Gly	Thr	Gln	Glu	Ser	Ile	Ser
	770					775					780				
Leu	Leu	Glu	Val	Ser	Thr	Leu	Gly	Lys	Ala	Lys	Thr	Glu	Pro	Asn	Lys
785					790					795					800
Cys	Val	Ser	Gln	Cys	Ala	Ala	Phe	Glu	Asn	Pro	Lys	Gly	Leu	Ile	His

A000206

5,747,282

85

86

-continued

805										810					815				
Gly	Cys	Ser	Lys	Asp	Asn	Arg	Asn	Asp	Thr	Glu	Gly	Phe	Lys	Tyr	Pro				
			820					825					830						
Leu	Gly	His	Glu	Val	Asn	His	Ser	Arg	Glu	Thr	Ser	Ile	Glu	Met	Glu				
		835					840					845							
Glu	Ser	Glu	Leu	Asp	Ala	Gln	Tyr	Leu	Gln	Asn	Thr	Phe	Lys	Val	Ser				
	850					855					860								
Lys	Arg	Gln	Ser	Phe	Ala	Pro	Phe	Ser	Asn	Pro	Gly	Asn	Ala	Glu	Glu				
865					870				875	875					880				
Glu	Cys	Ala	Thr	Phe	Ser	Ala	His	Ser	Gly	Ser	Leu	Lys	Lys	Gln	Ser				
				885					890					895					
Pro	Lys	Val	Thr	Phe	Glu	Cys	Glu	Gln	Lys	Glu	Glu	Asn	Gln	Gly	Lys				
			900					905					910						
Asn	Glu	Ser	Asn	Ile	Lys	Pro	Val	Gln	Thr	Val	Asn	Ile	Thr	Ala	Gly				
		915					920					925							
Phe	Pro	Val	Val	Gly	Gln	Lys	Asp	Lys	Pro	Val	Asp	Asn	Ala	Lys	Cys				
	930					935					940								
Ser	Ile	Lys	Gly	Gly	Ser	Arg	Phe	Cys	Leu	Ser	Ser	Gln	Phe	Arg	Gly				
945					950					955					960				
Asn	Glu	Thr	Gly	Leu	Ile	Thr	Pro	Asn	Lys	His	Gly	Leu	Leu	Gln	Asn				
			965						970					975					
Pro	Tyr	Arg	Ile	Pro	Pro	Leu	Phe	Pro	Ile	Lys	Ser	Phe	Val	Lys	Thr				
			980					985					990						
Lys	Cys	Lys	Lys	Asn	Leu	Leu	Glu	Glu	Asa	Phe	Glu	Glu	His	Ser	Met				
		995					1000					1005							
Ser	Pro	Glu	Arg	Glu	Met	Gly	Asn	Glu	Asn	Ile	Pro	Ser	Thr	Val	Ser				
	1010					1015					1020								
Thr	Ile	Ser	Arg	Asn	Asn	Ile	Arg	Glu	Asn	Val	Phe	Lys	Glu	Ala	Ser				
1025					1030						1035				1040				
Ser	Ser	Asn	Ile	Asn	Glu	Val	Gly	Ser	Ser	Thr	Asn	Glu	Val	Gly	Ser				
			1045						1050					1055					
Ser	Ile	Asn	Glu	Ile	Gly	Ser	Ser	Asp	Glu	Asn	Ile	Gln	Ala	Glu	Leu				
		1060						1065					1070						
Gly	Arg	Asn	Arg	Gly	Pro	Lys	Leu	Asn	Ala	Met	Leu	Arg	Leu	Gly	Val				
		1075					1080					1085							
Leu	Gln	Pro	Glu	Val	Tyr	Lys	Gln	Ser	Leu	Pro	Gly	Ser	Asn	Cys	Lys				
	1090					1095					1100								
His	Pro	Glu	Ile	Lys	Lys	Gln	Glu	Tyr	Glu	Glu	Val	Val	Gln	Thr	Val				
1105					1110						1115				1120				
Asn	Thr	Asp	Phe	Ser	Pro	Tyr	Leu	Ile	Ser	Asp	Asn	Leu	Glu	Gln	Pro				
			1125						1130					1135					
Met	Gly	Ser	Ser	His	Ala	Ser	Gln	Val	Cys	Ser	Glu	Thr	Pro	Asp	Asp				
			1140					1145					1150						
Leu	Leu	Asp	Asp	Gly	Glu	Ile	Lys	Glu	Asp	Thr	Ser	Phe	Ala	Glu	Asn				
		1155					1160					1165							
Asp	Ile	Lys	Glu	Ser	Ser	Ala	Val	Phe	Ser	Lys	Ser	Val	Gln	Lys	Gly				
	1170					1175					1180								
Glu	Leu	Ser	Arg	Ser	Pro	Ser	Pro	Phe	Thr	His	Thr	His	Leu	Ala	Gln				
1185					1190						1195				1200				
Gly	Tyr	Arg	Arg	Gly	Ala	Lys	Lys	Leu	Glu	Ser	Ser	Glu	Glu	Asn	Leu				
			1205						1210					1215					
Ser	Ser	Glu	Asp	Glu	Glu	Leu	Pro	Cys	Phe	Gln	His	Leu	Leu	Phe	Gly				
			1220					1225					1230						

A000207

5,747,282

87

88

-continued

Lys	Val	Asn	Asn	Ile	Pro	Ser	Gln	Ser	Thr	Arg	His	Ser	Thr	Val	Ala		
		1235					1240					1245					
Thr	Glu	Cys	Leu	Ser	Lys	Asn	Thr	Glu	Glu	Asn	Leu	Leu	Ser	Leu	Lys		
		1250				1255					1260						
Asn	Ser	Leu	Asn	Asp	Cys	Ser	Asn	Gln	Val	Ile	Leu	Ala	Lys	Ala	Ser		
		1265			1270					1275					1280		
Gln	Glu	His	His	Leu	Ser	Glu	Glu	Thr	Lys	Cys	Ser	Ala	Ser	Leu	Phe		
				1285					1290					1295			
Ser	Ser	Gln	Cys	Ser	Glu	Leu	Glu	Asp	Leu	Thr	Ala	Asn	Thr	Asn	Thr		
			1300					1305					1310				
Gln	Asp	Pro	Phe	Leu	Ile	Gly	Ser	Ser	Lys	Gln	Met	Arg	His	Gln	Ser		
		1315					1320					1325					
Glu	Ser	Gln	Gly	Val	Gly	Leu	Ser	Asp	Lys	Glu	Leu	Val	Ser	Asp	Asp		
		1330				1335					1340						
Glu	Glu	Arg	Gly	Thr	Gly	Leu	Glu	Glu	Asn	Asn	Gln	Glu	Glu	Gln	Ser		
		1345			1350				1355						1360		
Met	Asp	Ser	Asn	Leu	Gly	Glu	Ala	Ala	Ser	Gly	Cys	Glu	Ser	Glu	Thr		
			1365						1370					1375			
Ser	Val	Ser	Glu	Asp	Cys	Ser	Gly	Leu	Ser	Ser	Gln	Ser	Asp	Ile	Leu		
			1380					1385					1390				
Thr	Thr	Gln	Gln	Arg	Asp	Thr	Met	Gln	His	Asn	Leu	Ile	Lys	Leu	Gln		
		1395				1400						1405					
Gln	Glu	Met	Ala	Glu	Leu	Glu	Ala	Val	Leu	Glu	Gln	His	Gly	Ser	Gln		
		1410				1415					1420						
Pro	Ser	Asn	Ser	Tyr	Pro	Ser	Ile	Ile	Ser	Asp	Ser	Ser	Ala	Leu	Glu		
		1425			1430					1435					1440		
Asp	Leu	Arg	Asn	Pro	Glu	Gln	Ser	Thr	Ser	Glu	Lys	Ala	Val	Leu	Thr		
			1445						1450					1455			
Ser	Gln	Lys	Ser	Ser	Glu	Tyr	Pro	Ile	Ser	Gln	Asn	Pro	Glu	Gly	Leu		
			1460					1465					1470				
Ser	Ala	Asp	Lys	Phe	Glu	Val	Ser	Ala	Asp	Ser	Ser	Thr	Ser	Lys	Asn		
		1475					1480					1485					
Lys	Glu	Pro	Gly	Val	Glu	Arg	Ser	Ser	Pro	Ser	Lys	Cys	Pro	Ser	Leu		
		1490				1495					1500						
Asp	Asp	Arg	Trp	Tyr	Met	His	Ser	Cys	Ser	Gly	Ser	Leu	Gln	Asn	Arg		
		1505			1510					1515					1520		
Asn	Tyr	Pro	Ser	Gln	Glu	Glu	Leu	Ile	Lys	Val	Val	Asp	Val	Glu	Glu		
			1525						1530					1535			
Gln	Gln	Leu	Glu	Glu	Ser	Gly	Pro	His	Asp	Leu	Thr	Glu	Thr	Ser	Tyr		
		1540				1545						1550					
Leu	Pro	Arg	Gln	Asp	Leu	Glu	Gly	Thr	Pro	Tyr	Leu	Glu	Ser	Gly	Ile		
		1555				1560						1565					
Ser	Leu	Phe	Ser	Asp	Asp	Pro	Glu	Ser	Asp	Pro	Ser	Glu	Asp	Arg	Ala		
		1570				1575					1580						
Pro	Glu	Ser	Ala	Arg	Val	Gly	Asn	Ile	Pro	Ser	Ser	Thr	Ser	Ala	Leu		
		1585			1590					1595					1600		
Lys	Val	Pro	Gln	Leu	Lys	Val	Ala	Glu	Ser	Ala	Gln	Ser	Pro	Ala	Ala		
			1605					1610						1615			
Ala	His	Thr	Thr	Asp	Thr	Ala	Gly	Tyr	Asn	Ala	Met	Glu	Glu	Ser	Val		
		1620					1625						1630				
Ser	Arg	Glu	Lys	Pro	Glu	Leu	Thr	Ala	Ser	Thr	Glu	Arg	Val	Asn	Lys		
		1635				1640						1645					
Arg	Met	Ser	Met	Val	Val	Ser	Gly	Leu	Thr	Pro	Glu	Glu	Phe	Met	Leu		
		1650				1655					1660						

A000208

5,747,282

89

90

-continued

Val Tyr Lys Phe Ala Arg Lys His His Ile Thr Leu Thr Asn Leu Ile
1665 1670 1675 1680
Thr Glu Glu Thr Thr His Val Val Met Lys Thr Asp Ala Glu Phe Val
1685 1690 1695
Cys Glu Arg Thr Leu Lys Tyr Phe Leu Gly Ile Ala Gly Gly Lys Trp
1700 1705 1710
Val Val Ser Tyr Phe Trp Val Thr Gln Ser Ile Lys Glu Arg Lys Met
1715 1720 1725
Leu Asn Glu His Asp Phe Glu Val Arg Gly Asp Val Val Asn Gly Arg
1730 1735 1740
Asn His Gln Gly Pro Lys Arg Ala Arg Glu Ser Gln Asp Arg Lys Ile
1745 1750 1755 1760
Phe Arg Gly Leu Glu Ile Cys Cys Tyr Gly Pro Phe Thr Asn Met Pro
1765 1770 1775
Thr Asp Gln Leu Glu Trp Met Val Gln Leu Cys Gly Ala Ser Val Val
1780 1785 1790
Lys Glu Leu Ser Ser Phe Thr Leu Gly Thr Gly Val His Pro Ile Val
1795 1800 1805
Val Val Gln Pro Asp Ala Trp Thr Glu Asp Asn Gly Phe His Ala Ile
1810 1815 1820
Gly Gln Met Cys Glu Ala Pro Val Val Thr Arg Glu Trp Val Leu Asp
1825 1830 1835 1840
Ser Val Ala Leu Tyr Gln Cys Gln Glu Leu Asp Thr Tyr Leu Ile Pro
1845 1850 1855
Gln Ile Pro His Ser His Tyr
1860

(2) INFORMATION FOR SEQ ID NO:3:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (i i) MOLECULE TYPE: DNA (genomic)
- (i i i) HYPOTHETICAL: NO
- (v i) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo sapiens
- (v i i) IMMEDIATE SOURCE:
 - (B) CLONE: #754 A
- (x i) SEQUENCE DESCRIPTION: SEQ ID NO:3:

CTAGCCTGGG CAACAAACGA

20

(2) INFORMATION FOR SEQ ID NO:4:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (i i) MOLECULE TYPE: DNA (genomic)
- (i i i) HYPOTHETICAL: NO
- (v i) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo sapiens
- (v i i) IMMEDIATE SOURCE:

5,747,282

91

92

-continued

(B) CLONE: s754 B

(3 i) SEQUENCE DESCRIPTION: SEQ ID NO:4:

GCAGGAAGCA GGAATGGAAC

20

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens

(v i i) IMMEDIATE SOURCE:

(B) CLONE: s975 A

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:5:

TAGGAGATGG ATTATTGGTG

20

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens

(v i i) IMMEDIATE SOURCE:

(B) CLONE: 8975 B

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:6:

AGGCAACTTT GCAATGAGTG

20

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 22 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens

(v i i) IMMEDIATE SOURCE:

(B) CLONE: td1474 A

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:7:

CAGAGTGAGA CCTTGTCCTCA AA

22

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 23 base pairs

5,747,282

93

94

-continued

<div>(B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear</div> <div>(i i) MOLECULE TYPE: DNA (genomic)</div> <div>(i i i) HYPOTHETICAL: NO</div> <div>(v i) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens</div> <div>(v i i) IMMEDIATE SOURCE: (B) CLONE: tdj1474 B</div> <div>(x i) SEQUENCE DESCRIPTION: SEQ ID NO:8:</div> <div>T T C T G C A A A C A C C T T A A A C T C A G</div>		2 3
<div>(2) INFORMATION FOR SEQ ID NO:9:</div> <div>(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 20 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear</div> <div>(i i) MOLECULE TYPE: DNA (genomic)</div> <div>(i i i) HYPOTHETICAL: NO</div> <div>(v i) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens</div> <div>(v i i) IMMEDIATE SOURCE: (B) CLONE: tdj1239 A</div> <div>(x i) SEQUENCE DESCRIPTION: SEQ ID NO:9:</div> <div>A A C C T G G A A G G C A G A G G T T G</div>		2 0
<div>(2) INFORMATION FOR SEQ ID NO:10:</div> <div>(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 21 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear</div> <div>(i i) MOLECULE TYPE: DNA (genomic)</div> <div>(i i i) HYPOTHETICAL: NO</div> <div>(v i) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens</div> <div>(v i i) IMMEDIATE SOURCE: (B) CLONE: tdj1239 B</div> <div>(x i) SEQUENCE DESCRIPTION: SEQ ID NO:10:</div> <div>T C T G T A C C T G C T A A G C A G T G G</div>		2 1
<div>(2) INFORMATION FOR SEQ ID NO:11:</div> <div>(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 111 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: double (D) TOPOLOGY: linear</div> <div>(i i) MOLECULE TYPE: cDNA</div> <div>(i i i) HYPOTHETICAL: NO</div> <div>(v i) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens</div>		

5,747,282

95

96

-continued

(i x) FEATURE:	
(A) NAME/KEY: CDS	
(B) LOCATION: 2..111	
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:11:	
G GKC TTA CTC TGT TGT CCC AGC TGG AGT ACA GWG TGC GAT CAT GAG	46
Xaa Leu Leu Cys Cys Pro Ser Trp Ser Thr Xaa Cys Asp His Glu	
1865 1870 1875	
GCT TAC TGT TGC TTG ACT CCT AGG CTC AAG CGA TCC TAT CAC CTC AGT	94
Ala Tyr Cys Cys Leu Thr Pro Arg Leu Lys Arg Ser Tyr His Leu Ser	
1880 1885 1890 1895	
CTC CAA GTA GCT GGA CT	111
Leu Gln Val Ala Gly	
1900	
(2) INFORMATION FOR SEQ ID NO:12:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 36 amino acids	
(B) TYPE: amino acid	
(D) TOPOLOGY: linear	
(i i) MOLECULE TYPE: protein	
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:12:	
Xaa Leu Leu Cys Cys Pro Ser Trp Ser Thr Xaa Cys Asp His Glu Ala	
1 5 10 15	
Tyr Cys Cys Leu Thr Pro Arg Leu Lys Arg Ser Tyr His Leu Ser Leu	
20 25 30	
Gln Val Ala Gly	
35	
(2) INFORMATION FOR SEQ ID NO:13:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 1534 base pairs	
(B) TYPE: nucleic acid	
(C) STRANDEDNESS: double	
(D) TOPOLOGY: linear	
(i i) MOLECULE TYPE: DNA (genomic)	
(i i i) HYPOTHETICAL: NO	
(i v) ANTI-SENSE: NO	
(v i) ORIGINAL SOURCE:	
(A) ORGANISM: Homo sapiens	
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:13:	
GAGGCTAGAG GGCAGGCACT TTATGGCAAA CTCAGGTAGA ATTCTTCCTC TTCCGTCTCT	60
TTCTTTTAC GTCATCGGGG AGACTGGGTG GCAATCGCAG CCCGAGAGAC GCATGGCTCT	120
TTCTGCCCTC CATCCTCTGA TGTACCTTGA TTTTCGTATTC TGAGAGGCTG CTGCTTAGCG	180
GTAGCCCTT GGTITCCGTG GCAACGGAAA AGCGCGGGAA TTACAGATAA ATTAAAAC TG	240
CGACTGCGCG GCGTGAGCTC GCTGAGACTT CCTGGACCCC GCACCAGGCT GTGGGGTTTC	300
TCAGATAACT GGGCCCCCTGC GCTCAGGAGG CCTTCACCCCT CTGCTCTGGG TAAAGGTAGT	360
AGAGTCCCGG GAAAGGGACA GGGGGCCCAA GTGATGCTCT GGGGTACTGG CGTGGGAGAG	420
TGGATTTCG AAGCTGACAG ATGGGTATT TTTGACGGGG GGTAGGGGCG GAACCTGAGA	480
GGCGTAAGGC GTTGTGAACC CTGGGGAAGG GGGCAGTTTG TAGGTCGCGA GGGAAAGCGCT	540
GAGGATCAGG AAGGGGGCAC TGAGTGTCG TGGGGGAATC CTCGTGATAG GAACTGGAAT	600

5,747,282

97

98

-continued-

ATGCTTTGAG	GGGGACACTA	TGCTTTTAAA	AACGTCGGCT	GGTCATGAGG	TCAGGAGTTC	660
CAGACCAGCC	TGACCAACGT	GGTGAAACTC	CGTCTCTACT	AAAAATACNA	AAATTAGCCG	720
GCGGTGGTGC	CGCTCCAGCT	ACTCAGGAGG	CTGAGGCAGG	AGAATCGCTA	GAACCCGGGA	780
GGCGGAGGTT	GCAGTGAGCC	GAGATCGCGC	CATTGCACTC	CAGCCTGGGC	GACAGAGCGA	840
GACTGTCTCA	AAACAAAACA	AAACAAAACA	AAACAAAAAA	CACCGGCTGG	TATGTATGAG	900
AGGATGGGAC	CTTG TGGAAG	AAGAGGTGCC	AGGAATATGT	CTGGGAAGGG	GAGGAGACAG	960
GATTTTG TGG	GAGGGAGAAC	TTAAGAACTG	GATCCATTTG	CGCCATTGAG	AAAGCGCAAG	1020
AGGGAAAGTAG	AGGAGCGTCA	GTAGTAACAG	ATGCTGCCGG	CAGGGATGTG	CTTGAGGAGG	1080
ATCCAGAGAT	GAGAGCAGGT	CAC TGGGAAA	GGTTAGGGGC	GGGGAGGGCT	TGATTGGTGT	1140
TGGTTTG TGC	GTTGTTGATT	TTGGTTTTAT	GCAAGAAAAA	GAAAAACAACC	AGAAACATTG	1200
GAGAAAGCTA	AGGCTACCCAC	CACCTACCCG	GTCACTCACT	CCTCTGTAGC	TTTCTCTTTC	1260
TTGGAGAAAAG	GAAAAGACCC	AAGGGGTTGG	CAGCGATATG	TGAAAAAATT	CAGAATTTAT	1320
GTTGTCTAAT	TACAAAAGC	AAC TTCTAGA	ATCTTTAAAA	ATAAAGGACG	TTGTCAATTAG	1380
TTCTTCTGGT	TTGTATTATT	CTAAACCTT	CCAAATCTTC	AAATTTACTT	TATTTTAAAA	1440
TGATAAAATG	AAGTTGTCAT	TTTATAAAACC	TTTTAAAAAG	ATATATATAT	ATGTTTTTCT	1500
AAATGTGTAA	AGTTCAATTG	AACAGAAAAGA	AATG			1534

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1924 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(i v) ANTI-SENSE: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: *Homo sapiens*

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:14:

GAGGCTAGAG	GGCAGGCACT	TTATGGCAAA	CTCAGGTAGA	ATTCTTCCTC	TTCGCTCTCT	60
TTCCTTTTAC	GTCAICGGGG	AGACTGGGTG	GCAATCGCAG	CCCAGAGAGAC	GCATGGCTCT	120
TTCTGCCCTC	CATCCTCTGA	TGTACCTTGA	TTTCGTATTC	TGAGAGGGCTG	CTGCTTAGCG	180
GTAGCCCTT	GGTTTCCGTG	GCAACGGAAA	AGCGCGGGAA	TTACAGATAA	ATTAAAACTG	240
CGACTGCGCG	GCCTGAGCTC	GCTGAGACTT	CCTGGACCCC	GCACCAGGCT	GTGGGGTTTC	300
TCAGATAACT	GGGCCCTGCG	GCTCAGGAGG	CCTTCACCCT	CTGCTCTGGG	TAAAGGTAGT	360
AGAGTCCCGG	GAAAGGGACA	GGGGGCCCAA	GTGATGCTCT	GGGGTACTGG	CGTGGGAGAG	420
TGGATTTCGG	AAGCTGACAG	ATGGGTATTC	TTTGACGGGG	GGTAGGGGCG	GAACCTGAGA	480
GGCGTAAGGC	GTTGTGAACC	CTGGGGAGGG	GGGCAGTTTG	TAGGTCGCGA	GGGAAGCGCT	540
GAGGATCAGG	AAGGGGGCAC	TGAGTGTCCG	TGGGGGAATC	CTCGTGATAG	GAACCTGGAAT	600
ATGCCTTGAG	GGGGACACTA	TGTCTTTAAA	AACGTGGGCT	GGTCATGAGG	TCAGGAGTTC	660
CAGACCAGCC	TGACCAACGT	GGTGAAACTC	CGTCTCTACT	AAAAATACNA	AAATTAGCCG	720
GGCGTGGTGC	CGCTCCAGCT	ACTCAGGAGG	CTGAGGCAGG	AGAATCGCTA	GAACCCGGGA	780
GGCGGAGGTT	GCAGTGAGCC	GAGATCGCGC	CATTGCACTC	CAGCCTGGGC	GACAGAGCGA	840

99		5,747,282		100	
-continued					
GACTGTCTCA	AAACAAAACA	AAACAAAACA	AAACAAAAAA	CACCGGCTGG	TATGTATGAG 900
AGGATGGGAC	CTTGTGGAAG	AAGAGGTGCC	AGGAATATGT	CTGGGAAGGG	GAGGAGACAG 960
GATTTTGTGG	GAGGGAGAAC	TTAAGAACTG	GATCCATTTG	CGCCATTGAG	AAAGCGCAAG 1020
AGGGAAGTAG	AGGAGCGTCA	GTAGTAACAG	ATGCTGCCGG	CAGGGATGTG	CTTGAGGAGG 1080
ATCCAGAGAT	GAGAGCAGGT	CAC TGGGAAA	GGTTAGGGGC	GGGGAGGCCT	TGATTGGTGT 1140
TGGTTTGGTC	GTTGTTGATT	TTGGTTTTAT	GCAAGAAAAA	GAAAACAACC	AGAAACATTG 1200
GAGAAAAGCTA	AGGCTACCAC	CACCTACCCG	GTCAGTCACT	CCTCTGTAGC	TTTCTCTTTC 1260
TTGGAGAAAAG	GAAAAGACCC	AAGGGGTTGG	CAGCGATATG	TGAAAAAATT	CAGAAITTTAT 1320
GTTGTCTAAT	TACAAAAAGC	AACTTCTAGA	ATCTTTAAAA	ATAAAGGACG	TTGTCAATTAG 1380
TTCTTCTGGT	TTGTATTATT	CTAAAACCTT	CCAAATCTTC	AAATTTACTT	TATTTTAAAA 1440
TGATAAAATG	AAGTTGTCAT	TTTATAAACCC	TTTTAAAAAG	ATATATATAT	ATGTTTTTCT 1500
AATGTGTTAA	AGTTCATTGG	AACAGAAAGA	AATGGATTTA	TCTGCTCTTC	GCGTTGAAGA 1560
AGTACAAAAT	GTCATTAATG	CTATGCAGAA	AATCTTAGAG	TGTCCCATCT	GGTAAGTCAG 1620
CACAAGAGTG	TATTAATTTG	GGATTCCCTAT	GATTATCTCC	TATGCAAATG	AACAGAATTG 1680
ACCTTACATA	CTAGGGAAGA	AAAGACATGT	CTAGTAAGAT	TAGGCTATTG	TAATTGCTGA 1740
TTTTCTTAAC	TGAAGAACTT	TAAAAATATA	GAAAATGATT	CCTTGTCTC	CATCCACTCT 1800
GCCTCTCCCA	CTCCTCTCCT	TTTCAACACA	ATCCTGTGGT	CCGGGAAAAG	CAGGGCTCTG 1860
TCTTGATTGG	TTCTGCACTG	GGCAGGATCT	GTTAGATACT	GCATTTGCTT	TCTCCAGCTC 1920
TAAA					1924

(2) INFORMATION FOR SEQ ID NO:15:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 631 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(i v) ANTI-SENSE: NO

(v i) ORIGINAL SOURCE:

- (A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:15:

AAATGCTGAT	GATAGTATAG	AGTATTGAAG	GGATCAATAT	AATTCTGTTT	TGATAICTGA 60
AAGCTCACTG	AAGGTAAGGA	TCGTATTCTC	TGCTGTATTTC	TCAGTTCCTG	ACACAGCAGA 120
CATTTAATAA	ATATTGAACG	AAC TTGAGGC	CTTATGTTGA	CTCAGTCATA	ACAGCTCAAA 180
GTTGAAC TTA	TTCACTAAGA	ATAGCTTTAT	TTTTAAATAA	ATTATTGAGC	CTCATTTATT 240
TTCTTTTTCT	CCCCCCCCCTA	CCCTGCTAGT	CTGGAGTTGA	TCAAGGAACC	TGTCTCCACA 300
AAGTGTGACC	ACATATTTTG	CAAGTAAGTT	TGAATGTGTT	ATGTGGCTCC	ATTATTAGCT 360
TTTGTTTTTG	TCCTTCATAA	CCCAGGAAAC	ACCTAAC TTT	ATAGAAGCTT	TACTTCTTTC 420
AATTAAGTGA	GAACGAAAAAT	CCA ACTCCAT	TTCATTCTTT	CTCAGAGAGT	ATATAGTTAT 480
CAAAAAGTTGG	TTGTAATCAT	AGTTCTCTGGT	AAAGTTTTGA	CATATATTAT	CTTTTTTTTT 540
TTTTGAGACA	AGTCTCGCTC	TGTCGCCCCAG	GCTGGAGTGC	AGTGGCATGA	GGCTTGCTCA 600
CTGCACCTCC	GCCCCCGAGT	TCAGCGACTC	T		631

5,747,282

101

102

-continued

(2) INFORMATION FOR SEQ ID NO:16:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 481 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(i v) ANTI-SENSE: NO

(v i) ORIGINAL SOURCE:

- (A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:16:

TGAGATCTAG	ACCACATGGT	CAAAGAGATA	GAATGTGAGC	AATAAATGAA	CCTTAAATTT	60
TTCAACAGCT	ACTTTTTTTT	TTTTTTTTTG	AGACAGGGKC	TTACTCTGTT	GTCCCAGCTG	120
GAGTACAGWG	TGCGATCATG	AGGCTTACTG	TTGCTTGACT	CCTAGGCTCA	AGCGATCCTA	180
TCACCTCAGT	CTCCAAGTAG	CTGGACTGTA	AGTGCACACC	ACCATATCCA	GCTAAATTTT	240
GTGTTTTCTG	TAGAGACGGG	GTTTCGCCAT	GTTTCCCAGG	CTGGTCTTGA	ACTTTGGGCT	300
TAACCCGTCT	GCCCACCTAG	GCATCCCCAA	GTGCTAGGAT	TACAGGTGTG	AGTCATCATG	360
CCTGGCCAGT	ATTTTAGTTA	GCTCTGTCTT	TTCAAGTCAT	ATACAAAGTT	ATTTTCTTTT	420
AAGTTTAGTT	AACAACCTTA	TATCATGTAT	TCTTTTCTAG	CATAAAGAAA	GATTTCGAGGC	480
C						481

(2) INFORMATION FOR SEQ ID NO:17:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 522 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(i v) ANTI-SENSE: NO

(v i) ORIGINAL SOURCE:

- (A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:17:

TGTGATCATA	ACAGTAAGCC	ATATGCAATG	AAGTTCAGTT	TTCATAGATC	ATTGCTTATG	60
TAGTTTAGGT	TTTTGCTTAT	GCAGCATCCA	AAAACAATTA	GGAAACTATT	GCTTGTAATT	120
CACCTGCCAT	TACTTTTTAA	ATGGCTCTTA	AGGGCAGTTG	TGAGATTATC	TTTTCATGGC	180
TATTTGCCTT	TTGAGTATTC	TTTCTACAAA	AGGAAGTAAA	TTAAATTGTT	CTTTCCTTCT	240
TTATAATTTA	TAGATTTTGC	ATGCTGAAAC	TTCTCAACCA	GAAGAAAGGG	CCTTCACAGT	300
GTCTTTTATG	TAAGAATGAT	ATAACCAAAA	GGTATATAAT	TTGGTAATGA	TGCTAGGTTG	360
GAAGCAACCA	CAGTAGGAAA	AAGTAGAAAT	TATTTAATAA	CATAGCGTTC	CTATAAAACC	420
ATTTCATCAGA	AAAATTTATA	AAAGAGTTTT	TAGCACACAG	TAAATTATTT	CCAAAGTTAT	480
TTTCCTGAAA	GTTTTATGGG	CATCTGCCTT	ATACAGGTAT	TG		522

(2) INFORMATION FOR SEQ ID NO:18:

- (i) SEQUENCE CHARACTERISTICS:

5,747,282

103

104

-continued

(A) LENGTH: 465 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(i v) ANTI-SENSE: NO

(v i) ORIGINAL SOURCE:

(A) ORGANISM: *Homo sapiens*

(k i) SEQUENCE DESCRIPTION: SEQ ID NO:18:

GGTAGGCTTA	AATGAATGAC	AAAAAGTTAC	TAAATCACTG	CCATCACACG	GTTTATACAG	60
ATGTCAATGA	TGTATTGATT	ATAGAGGTTT	TCTACTGTTG	CTGCATCTTA	TTTTTATTTG	120
TTTACATGTC	TTTTCTTATT	TTAGTGTCCT	TAAAAGGTTG	ATAATCACTT	GCTGAGTGTG	180
TTTCTCAAAC	AATTTAATTT	CAGGAGCCTA	CAAGAAAGTA	CGAGATTTAG	TCAACTTGTT	240
GAAGAGCTAT	TGAAAAATCAT	TTGTGCTTTT	CAGCTTGACA	CAGGTTTGGG	GTGTAAGTGT	300
TGAATATCCC	AAGAATGACA	CTCAAGTGCT	GTCCATGAAA	ACTCAGGAAG	TTTGCACAAT	360
TACTTTCTAT	GACGTGGTGA	TAAGACCTTT	TAGTCTAGGT	TAATTTTAGT	TCTGTATCTG	420
TAACTATTTT	TAAAAAATTA	CTCCCACTGG	TCACACACCT	TATTT		465

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 513 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(i v) ANTI-SENSE: NO

(v i) ORIGINAL SOURCE:

(A) ORGANISM: *Homo sapiens*

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:19:

AAAAAATCAC	AGGTAACCTT	AATGCATTGT	CTTAACACAA	CAAAGAGCAT	ACATAGGGTT	60
TCTCTTGGTT	TCTTTGATTA	TAATTCATAC	ATTTTTCTCT	AAC TGCAAA C	ATAATGTTTT	120
CCCTTGTATT	TTACAGATGC	AAACAGCTAT	AATTTTGCAA	AAAAGGAAAA	TAAC TCTCCT	180
GAACATCTAA	AAAGATGAAGT	TTCTATCATC	CAAAGTATGG	GCTACAGAAA	CCGTGCCAAA	240
AGACTTCTAC	AGAGTGAACC	CGAAAAATCCT	TCCITGGTAA	AACCATTTGT	TTTCTTCTTC	300
TTCTTCTTCT	TCTTTTCTTT	TTTTTTTCTT	TTTTTTTTTG	AGATGGAGTC	TTGCTCTGTG	360
GCCCAGGCTA	GAAGCAGTCC	TCCTGCCTTA	GCCNCCTTAG	TAGCTGGGAT	TACAGGCACG	420
CGCACCATGC	CAGGCTAATT	TTTGTAATTT	TAGTAGAGAC	GGGGTTTCAT	CATGTTGGCC	480
AGGCTGGTCT	CGAACTCCTA	ACCTCAGGTG	ATC			513

(2) INFORMATION FOR SEO ID NO:20:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 6769 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

5,747,282

105

106

-continued

(i i i)	HYPOTHETICAL:	NO	
(i v)	ANTI-SENSE:	NO	
(v i)	ORIGINAL SOURCE:		
(A)	ORGANISM:	Homo sapiens	
(x i)	SEQUENCE DESCRIPTION:	SEQ ID NO:20:	
ATGATGGAGA	TCTTAAAAAG	TAATCATTTCT	GGGGCTGGGC GTAGTAGCTT GCACCTGTAA 60
TCCCAGCACT	TCGGGAGGCT	GAGGCAGGCA	GATAATTTGA GGTCAGGAGT TTGAGACCAG 120
CCTGGCCAAC	ATGGTGAAAC	CCATCTCTAC	TAAAAATACA AAAATTAGCT GGGTGTGGTG 180
GCACGTACCT	GTAATCCCAG	CTACTCGGGA	GGCGGAAGCA CAAGAATTGC TTGAACCTAG 240
GACGCGGAGG	TTGCAGCGAG	CCAAGATCGC	GCCACTGCAC TCCAGCCTGG GCCGTAGAGT 300
GAGACTCTGT	CTCAAAAAAG	AAAAAAAAGT	AATTGTICTA GCTGGGCGCA GTGGCTCTTG 360
CCTGTAATCC	CAGCACTTTG	GGAGGCCAAG	GCGGGTGGAT CTCGAGTCCT AGAGTTCAAG 420
ACCAGCCTAG	GCAATGTGGT	GAAACCCCAT	CGCTACAAAA AATACAAAAA TTAGCCAGGC 480
ATGGTGGCGT	GCGCATGTAG	TCCCAGCTCC	TTGGGAGGCT GAGGTGGGAG GATCACTTGA 540
ACCCAGGAGA	CAGAGGTTGC	AGTGAACCGA	GATCAGCCCA CCACGCTCCA GCCTGGGCAA 600
CAGAACAAGA	CTCTGTCTAA	AAAAATACAA	ATAAAATAAA AGTAGTTCTC ACAGTACCAG 660
CATTCATTTT	TCAAAAGATA	TAGAGCTAAA	AAGGAAGGAA AAAAAAAGTA ATGTTGGGCT 720
TTTAAATACT	CGTTCCCTATA	CTAAATGTTC	TTAGGAGTGC TGGGGTTTTA TTGTCATCAT 780
TTATCCTTTT	TAAAAATGTT	ATTGGCCAGG	CACGGTGGCT CATGGCTGTA ATCCCAGCAC 840
TTTGGGAGGC	CGAGGCAGGC	AGATCACCTG	AGGTCAGGAG TGTGAGACCA GCCTGGCCAA 900
CATGGCGAAA	CCTGTCTCTA	CTAAAAATAC	AAAAATTAACTAGGCGTGGT GGTGTACGCC 960
TGTAGTCCCA	GCTACTCGGG	AGGCTGAGGC	AGGAGAATCA ACTGAACCAAG GGAGGTGGAG 1020
GTTCAGTGT	GCCGAGATCA	CGCCACTGCA	CTCTAGCCTG GCAACAGAGC AAGATTCTGT 1080
CTCAAAAAAA	AAAAACATAT	ATACACATAT	ATCCCAAAAGT GCTGGGATTA CATATATATA 1140
TATATATATA	TATTATATAT	ATATATATAT	ATATATGTGA TATATATGTG ATATATATAT 1200
AACATATATA	TATGTAATAT	ATAATGTGATA	TATATATAAT ATATATATGT AATATATATG 1260
TGATATATAT	ATATACACAC	ACACACACAT	ATATATGTAT GTGTGTGTAC ACACACACAC 1320
ACAAATTAGC	CAGGCATAGT	TGCACACGCT	TGGTAGACCC AGCTACTCAG GAGGCTGAGG 1380
GAGGAGAATC	TCTTGAACCT	AGGAGGCGGA	GOTTGCAGTG AGCTGAGATT GCGCCACTGC 1440
ACTCCAGCCT	GGGTGACAGA	GCAGGACTCT	GTACACCCCC CAAAAACAAA AAAAAAGTTA 1500
TCAGATGTGA	TTGGAATGTA	TATCAAGTAT	CAGCTTCAAA ATATGCTATA TTAATACTTC 1560
AAAAATTACA	CAAATAATAC	ATAATCAGGT	TTGAAAAATT TAAGACAACM SAARAAAAAA 1620
WY CMAATCAC	AMATATCCCA	CACATTTTAT	TATIMCTMCT MCWATTATTT TGWAGAGMCT 1680
GGGTCTCAC	Y C Y KTTGCTWA	TGCTG-	
GTCTT	TGAAC Y CCY K	GCCY CAARCA	RTCCTSC TCC 1740
ABCCTCCCAA	RGTGCTGGGG	ATWATAGGCA	TGARCTAACC GCACCCAGCC CCAGACATTT 1800
TAGTGTGTAA	ATTCTGCGG	ATTTTTTCAA	GGCATCATAC ATGTTAGCTG ACTGATGATG 1860
GTCAATTTAT	TTTGTCCTAG	GTGTCAAGTT	TCTCTTCAGG AGGAAAAAGCA CAGAACTGGC 1920
CAACAATTGC	TTGACTGTTC	TTTACCATAC	TGTTTAGCAG GAAACCAAGTC TCAGTGTCCA 1980
ACTCTCTAAC	CTTGGAACCT	TGAGAACTCT	GAGGACAAAAG CAGCGGATAC AACCTCAAAA 2040
GACGTCTGTC	TACATTGAAT	TGGGTAAGGG	TCTCAGGTTT TTTAAGTATT TAATAATAAT 2100

5,747,282

107

108

-continued

TGCTGGATTTC	CTTATCTTAT	AGTTTTTGCCA	AAAACTTTGG	TCATAATTTG	TATTTGTGGT	2160
AGGCAGCTTT	GGGAAGTGAA	TTTTATGAGC	CCTATGGTGA	GTATATAAAA	ATGTAAAAGA	2220
CGCAGTTCCC	ACCTTGAAGA	ATCTTACTTT	AAAAAGGGAG	CAAAAGAGGC	CAGGCATGGT	2280
GGCTCACACC	TGTAATCCCA	GCACTTTGGG	AGGCCAAAGT	GGGTGGATCA	CCTGAGGTCG	2340
GGAGTTCGAG	ACCAGCCTAG	CCAACATGGA	GAAACTCTGT	CTGTACCAA	AAATAAAAAA	2400
TTAGCCAGGT	GTGOTGGCAC	ATAACTGTAA	TCCCAGCTAC	TCGGGAGGCT	GAGGCAGGAG	2460
AATCACTTGA	ACCCGGGAGG	TGGAGGTTGC	GGTGAACCGA	GATCGCACCA	TTGCACTCCA	2520
GCCTGGGCAA	AAATAGCGAA	ACTCCATCTA	AAAAAAAAAA	AGAGAGCAAA	AGAAAGAMTM	2580
TCTGGTTTTA	AMTMGTGT	AATATGTTTT	TGGAAAGATG	GAGAGTAGCA	ATAAGAAAAA	2640
ACATGATGGA	TTGCTACAGT	ATTTAGTTCC	AAGATAAATT	GTACTAGATG	AGGAAGCCTT	2700
TTAAGAAGAG	CTGAATTGCC	AGGCGCAGTG	GCTCACGCCT	GTAATCCCA	CACTTTGGGA	2760
GGCCGAGGTG	GGCGGATCAC	CTGAGGTCGG	GAGTTCAAGA	CCAGCCTGAC	CAACATGGAG	2820
AAACCCCATC	TCTACTAAAA	AAAAAAAAAA	AAAAATTAGC	CGGGGTGGTG	GCTTATGCCT	2880
GTAATCCCA	CTACTCAGGA	GGCTGAGGCA	GGAGAATCGC	TTGAACCCAG	GAAGCAGAGG	2940
TTGCAGTGAG	CCAAGATCGC	ACCATTGCAC	TCCAGCCTAG	GCAACAAGAG	TGAAACTCCA	3000
TCTCAAAAAA	AAAAAAAAAAG	AGCTGAATCT	TGGCTGGGCA	GGATGGCTCG	TGCTGTAAAT	3060
CCTAACGCTT	TGGAAGACCG	AGGCAGAAGG	ATTGGTTGAG	TCCACGAGTT	TAAGACCAGC	3120
CTGGCCAACA	TAGGGGAACC	CTGTCTCTAT	TTTTAAAAAT	ATAATACATT	TTTGGCCGGT	3180
GCGGTGGCTC	ATGCCTGTAA	TCCCAATACT	TTGGGAGGCT	GAGGCAGGTA	GATCACCTGA	3240
GGTCAGAGTT	CGAGACCAGC	CTGGATAACC	TGGTGAAACC	CCTCTTTACT	AAAAATACAA	3300
AAAAAAAAAA	AAATTAGCTG	GGTGTGGTAG	CACATGCTTG	TAATCCCA	TACTTGGGAG	3360
GCTGAGGCAG	GAGAATCGCT	TGAACCAGGG	AGGCGGAGGT	TACAATGAGC	CAACACTACA	3420
CCACTGCACT	CCAGCCTGGG	CAATAGAGTG	AGACTGCATC	TCAAAAAAAT	AATAATTTTT	3480
AAAAATAATA	AATTTTTTTT	AGCTTATAAA	AAGAAAAGTT	GAGGCCAGCA	TAGTAGCTCA	3540
CATCTGTAAAT	CTCAGCAGTG	GCAGAGGATT	GCTTGAAAGCC	AGGAGTTTGA	GACCAGCCTG	3600
GGCAACATAG	CAAGACCTCA	TCTCTACAAA	AAAAATTTCT	TTTTAAATTA	GCTGGGTGTG	3660
GTGGTGTGCA	TCTGTAGTCC	CAGCTACTCA	GGAGGCAGAG	GTGAGTGGAT	ACATTGAACC	3720
CAGGAGTTTG	AGGCTGTAGT	GAGCTATGAT	CATGCCACTG	CACTCCAACC	TGGGTGACAG	3780
AGCAAGACCT	CAAAAAAA	AAAAAAAAAGA	GCTGCTGAGC	TCAGAATTCA	AACTGGGCTC	3840
TCAAATTGGA	TTTTCTTTT	GAATATATTT	ATAATTAAAA	AGGATAGCCA	TCTTTTGAGC	3900
TCCCAGGCAC	CACCATCTAT	TTATCATAAC	ACTTACTGTT	TTCCCCCTT	ATGATCATAA	3960
ATTCTTAGAC	AACAGGCATT	GTAAAAATAG	TTATAGTAGT	TGATATTTAG	GAGCACTTAA	4020
CTATAITCCA	GGCACTATTG	TGCTTTTCTT	GTATAACTCA	TTAGATGCTT	GTGAGACCTC	4080
TGAGATTGTT	CCTATTATAC	TTATTTTACA	GATGAGAAAA	TTAAGGCACA	GAGAAGTTAT	4140
GAAATTTTTT	CAAGGTATTA	AACCTAGTAA	GTGGCTGAGC	CATGATTCAA	ACCTAGGAAG	4200
TTAGATGTCA	GAGCCTGTGC	TTTTTTTTTT	TTTTTGTGTT	TGTTTTCAGT	AGAAACGGGG	4260
GTCTCACTTT	GTTGGCCAGG	CTGGTCTTGA	ACTCCTAACC	TCAAATAATC	CACCCATCTC	4320
GGCTCCTCA	AGTGCTGGGA	TTACAGGTGA	GAGCCACTGT	GCCTGGCGAA	GCCCATGCCT	4380
TTAACCACCT	CTCTGTATTA	CATACTAGCT	TAACTAGCAT	TGTACCTGCC	ACAGTAGATG	4440
CTCAGTAAAT	ATTTCTAGTT	GAATATCTGT	TTTCAACAA	GTACATTTTT	TTAACCTTTT	4500

A000218

109			5,747,282		110		
-continued							
TAATTAAGAA	AACTTTTATT	GATTTATTTT	TTGGGGGGGAA	ATTTTTTAGG	ATCTGATTCT	4560	
TCTGAAGATA	CCGTTAATAA	GGCAACTTAT	TGCAGGTGAG	TCAAAGAGAA	CCTTTGTCTA	4620	
TGAAGCTGGT	ATTTTCCTAT	TTAGTTAATA	TAAAGGATTG	ATGTTTCICT	CTTTTTAAAA	4680	
ATATTTTAAC	TTTTATTTTA	GGTTCAGGGA	TGTATGTGCA	GTTTGTTATA	TAGGTAAACA	4740	
CACGACTTGG	GATTTGGTGT	ATAGATTTTT	TTCATCATCC	GGGTACTAAG	CATACCCAC	4800	
AGTTTTTGT	TTGCTTTCCT	TCTGAATTTT	TCCCTCTTCC	CACCTTCCTC	CCTCAAGTAG	4860	
GCTGGTGTTT	CTCCAGACTA	GAATCATGGT	ATTGGAAGAA	ACCTTAGAGA	TCATCTAGTT	4920	
TAGTTCCTCT	ATTTTATAGT	GGAGGAAATA	CCCTTTTTGT	TTGTTGGATT	TAGTTATTAG	4980	
CACGTGTCAA	AGGAATTTAG	GATAACAGTA	GAACTCTGCA	CATGCTTGCT	TCTAGCAGAT	5040	
TGTTCTCTAA	GTTCCCTCATA	TACAGTAATA	TTGACACAGC	AGTAATTGTG	ACTGATGAAA	5100	
ATGTTCAAGG	ACTTCATTTT	CAACTCTTTC	TTTCCCTCTG	TCCTTATTTT	CACATATCTC	5160	
TCAAGCTTTG	TCTGTATGTT	ATATAATAAA	CTACAAGCAA	CCCCAACTAT	GTTACCTACC	5220	
TTCCTTAGGA	ATTATTGCTT	GACCCAGGTT	TTTTTTTTTT	TTTTTTTGGG	GACGGGGTCT	5280	
TGCCCTGTTG	CCAGGATGGA	GTGTAAGTGG	GCCATCTCGG	CTCACTGCAA	TCTCCAACTC	5340	
CCTGGTTCAA	GCGATTCTCC	TGTTCTCAATC	TCACGAGTAG	CTGGGACTAC	AGGTATACAC	5400	
CACCACGCCC	GGTTAATTGA	CCATTCCATT	TCITTCITTC	TCTCTTTTTT	TTTTTTTTTT	5460	
TTGAGACAGA	GTCTTGCTCT	GTTGCCCAGG	CTGGAGTACA	GAGGTGTGAT	CTCACCTCTC	5520	
CGCAACGTCT	GCCTCCCAGG	TTGAAGCCAT	ACTCCTGCCT	CAGCCTCTCT	AGTAGCTGGG	5580	
ACTACAGGCG	GCGGCCACCA	CACCCGGCTA	ATTTTTGTAT	TTTTAGTAGA	GATGGGGTTT	5640	
CACCATGTTG	GCCAGGCTGG	TCTTGAATCT	ATGACCTCAA	GTGGTCCACC	CGCCTCAGCC	5700	
TCCCAAAGTG	CTGGAATTAC	AGGCTTGAGC	CACCGTGCCC	AGCAACCATT	TCATTTCAAC	5760	
TAGAAGTTTC	TAAAGGAGAG	AGCAGCTTTC	ACTAACTAAA	TAAAGATTGGT	CAGCTTTCTG	5820	
TAATCGAAAG	AGCTAAAATG	TTTGATCTTG	GTCAATTTGAC	AGTTCTGCAT	ACATGTAACT	5880	
AGTGTTTCTT	ATTAGGACTC	TGTTCTTTTCC	CTATAGTGTG	GGAGATCAAG	AATTGTTACA	5940	
AATCACCCCT	CAAGGAACCA	GGGATGAAAT	CAGTTTGGAT	TCTGCAAAAA	AGGGTAATGG	6000	
CAAAGTTTGC	CAACTTAACA	GGCACTGAAA	AGAGAGTGGG	TAGATACAGT	ACTGTAATTA	6060	
GATTATTTCT	AAGACCATTT	GGGACCTTTA	CAACCCACAA	AATCTCTTGG	CAGAGTTAGA	6120	
GTATCATTTCT	CTGTCAAATG	TCGTGGTATG	GTCTGATAGA	TTTAAATGGT	ACTAGACTAA	6180	
TGTACCTATA	ATAAGACCTT	CTTGTAACCT	ATTGTTGCC	TTTCGCTTTT	TTTTTTGTTT	6240	
GTTTGTGTTG	TTTTTTTTTG	GATGGGGTCT	CACCTCTGTT	CCCAGGCTGG	AGTGCAGTGA	6300	
TGCAATCTTG	GCTCACTGCA	ACCTCCACCT	CCAAAGGCTC	AAGCTATCCT	CCCACCTCAG	6360	
CCTCCTGAGT	AGCTGGGACT	ACAGGCGCAT	GCCACCACAC	CCGGTTAATT	TTTTGTGGTT	6420	
TTATAGAGAT	GGGGTTTCAC	CATGTTACCG	AGGCTGGTCT	CAAACCTCTG	GACTCAAGCA	6480	
GTCTGCCAC	TTCAGCCTCC	CAAAGTGCTG	CAGTTACAGG	CTTGAGCCAC	TGTGCCTGGC	6540	
CTGCCCTTTA	CTTTTAAATT	GTGTATTTGT	GTTTCATCTT	TTACCTACTG	GTTTTTAAAT	6600	
ATAGGGAAGT	GTAAGTCTGT	AGATAGAACA	GAGTATTAAG	TAGACTTAAT	GGCCAGTAAT	6660	
CTTTAGAGTA	CATCAGAACC	AGTTTTCTGA	TGGCCAATCT	GCTTTTAAAT	CACCTTTAGA	6720	
CGTTAGAGAA	ATAGGTGTGG	TTTCTGCATA	GGGAAAATTC	TGAAATTAA		6769	

(2) INFORMATION FOR SEQ ID NO:21:

5,747,282

111

112

-continued

(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 4249 base pairs	
(B) TYPE: nucleic acid	
(C) STRANDEDNESS: double	
(D) TOPOLOGY: linear	
(i i) MOLECULE TYPE: DNA (genomic)	
(i i i) HYPOTHETICAL: NO	
(i v) ANTI-SENSE: NO	
(v i) ORIGINAL SOURCE:	
(A) ORGANISM: Homo sapiens	
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:21:	
GATCCTAAGT GGAAATAATC TAGGTAAATA GGAATTAAAT GAAAGAGTAT GAGCTACATC	60
TTCAGTATAC TTGGTAGTTT ATGAGGTTAG TTCTCTAAT ATAGCCAGTT GGTTGATTTT	120
CACCTCCAAG GTGTAIGAAG TATGIATTTT TTAATGACA ATTCAAGTTT TGAGTACCTT	180
GTTATTTTTT TATATTTTCA GCTGCTTG TG AATTTTCTGA GACGGATGTA ACAAATACTG	240
AACATCATCA ACCCAGTAAT AATGATTGTA ACACCACTGA GAAGCGTGCA GCTGAGAGGC	300
ATCCAGAAAA GTATCAGGGT AGTTCTGTTT CAAACTTGCA TGTGGAGCCA TGTGGCACAA	360
ATACTCATGC CAGCTCATT AAGCATGAGA ACAGCAGTTT ATTACTCACT AAAGACAGAA	420
TGAATGTAGA AAAGGCTGAA TTCTGTAATA AAAGCAAACA GCCTGGCTTA GCAAGGAAGC	480
AACATAACAG ATGGGCTGGA AGTAAGGAAA CATGTAATGA TAGGCGGACT CCCAGCACAG	540
AAAAAAGGT AGATCTGAAT GCTGATCCCC TGTGTGAGAG AAAAGAATGG AATAAGCAGA	600
AACTGCCATG CTCAGAGAAT CCTAGAGATA CTGAAGATGT TCCTTGGATA AACTAAATA	660
GCAGCATTCA GAAAGTTAAT GAGTGGTTTT CCAGAAGTGA TGAAGTGTTA GGTTCGTATG	720
ACTCACATGA TGGGGAGTCT GAATCAAATG CCAAAGTAGC TGATGTATTG GACGTTCTAA	780
ATGAGGTAGA TGAATATTCT GGTTCCTCAG AGAAAATAGA CTTACTGGCC AGTGATCCTC	840
ATGAGGCTTT AATATGTAAA AGTGAAAGAG TTCACTCCAA ATCAGTAGAG AGTAATATTG	900
AAGGCCAAAT ATTTGGGAAA ACCTATCGGA AGAAGGCAAG CCTCCCCAAC TTAAGCCATG	960
TAAGTGAAAA TCTAATTATA GGAGCATTTG TTAGTGAGCC ACAGATAATA CAAGAGCGTC	1020
CCCTCACAAA TAAATTAAG CGTAAAAGGA GACCTACATC AGGCCTTCAT CCTGAGGATT	1080
TTATCAAGAA AGCAGATTTG GCAGTTCAAA AGACTCCTGA AATGATAAAT CAGGGAAC TA	1140
ACCAAACGGA GCAGAATGGT CAAGTGATGA ATATTACTAA TAGTGGTCAT GAGAATAAAA	1200
CAAAAGGTGA TTCTATTGAG AATGAGAAAA ATCCTAACCC AATAGAATCA CTCGAAAAAG	1260
AATCTGCTTT CAAAACGAAA GCTGAACCTA TAAGCAGCAG TATAAGCAAT ATGGAACCTG	1320
AATTAAATAT CCACAATTCA AAAGCACCTA AAAAGAATAG GCTGAGGAGG AAGTCTTCTA	1380
CCAGGCATAT TCATGCGCTT GAACTAGTAG TCAGTAGAAA TCTAAGCCCA CCTAATTGTA	1440
CTGAATTGCA AATTGATAGT GTTCTAGCA GTGAAGAGAT AAAGAAAAAA AAGTACAACC	1500
AAATGCCAGT CAGGCACAGC AGAAACCTAC AACTCATGGA AGGTAAAGAA CCTGCAACTG	1560
GAGCCAAGAA GAGTAACAAG CCAAAATGAAC AGACAAGTAA AAGACATGAC AGCGATACTT	1620
TCCCAGAGCT GAAGTTAACA AATGCACCTG GTTCTTTTAC TAAGTGTTCA AATACCACTG	1680
AACTTAAAGA ATTTGTCAAT CCTAGCCTTC CAAGAGAAGA AAAAGAAGAG AACTAGAAAC	1740
AGTTAAAGTG TCTAATAATG CTGAAGACCC CAAAGATCTC ATGTTAAGTG GAGAAAGGGT	1800
TTTGCAAAC T GAAAGATCTG TAGAGAGTAG CAGTATTTCA TTGGTACCTG GTACTGATTA	1860
TGGCACTCAG GAAAGTATCT CGTACTGGA AGTTAGCACT CTAGGGAAGG CAAAAACAGA	1920

113			5,747,282		114		
						-continued	
ACCAAATAAA	TGTGTGAGTC	AGTGTGCAGC	ATTTGAAAAC	CCCAAGGGAC	TAATTCATGG	1980	
TTGTTCCAAA	GATAATAGAA	ATGACACAGA	AGGCTTTAAG	TATCCATTGG	GACATGAAGT	2040	
TAACCACAGT	CGGGAAACAA	GCATAGAAAT	GGAAGAAAGT	GAAC TTGATG	CTCAGTATTT	2100	
GCAGAATACA	TTCAAGGTTT	CAAAGCGCCA	GTCA TTGCT	CCGTTTTC	AAATCCAGGAAA	2160	
TGCAGAAGAG	GAATGTGCAA	CATTCTCTGC	CCACTCTGGG	TCCTTAAAGA	AACAAAGTCC	2220	
AAAAGTCACT	TTTGAATGTG	AACAAAAGGA	AGAAAAATCA	GGAAAGAAATG	AGTCTAATAT	2280	
CAAGCCTGTA	CAGACAGTTA	ATATCACTGC	AGGCTTTCCT	GTGGTTGGTC	AGAAAGATAA	2340	
GCCAGTTGAT	AATGCCAAAT	GTAGTATCAA	AGGAGGCTCT	AGGTTTTGTG	TATCATCTCA	2400	
GTTCAGAGGC	AACGAAACTG	GACTCATTAC	TCCAAATAAA	CATGGACTTT	TACAAAACCC	2460	
ATATCGTATA	CCACCACCTT	TTCCCATCAA	GTCA TTGTT	AAAACTAAAT	GTAAGAAAAA	2520	
TCTGCTAGAG	GAAAACITTG	AGGAACATTC	AATGTCACCT	GAAAGAGAAA	TGGGAAATGA	2580	
GAACATTCCA	AGTACAGTGA	GCACAATTAG	CCGTAATAAC	ATTAGAGAAA	ATGTTTTTAA	2640	
AGAAGCCAGC	TCAAGCAATA	TTAATGAAGT	AGGTTCCAGT	ACTAATGAAG	TGGGCTCCAG	2700	
TATTAATGAA	ATAGGTTCCA	GTGATGAAAA	CATTCAAGCA	GAAGTAGGTA	GAAACAGAGG	2760	
GCCAAAATTG	AATGCTATGC	TTAGATTAGG	GGTTTTGCAA	CCTGAGGTCT	ATAAACAAAG	2820	
TCTTCTGGA	AGTAATTGTA	AGCATCCTGA	AATAAAAAAG	CAAGAATATG	AAGAAGTAGT	2880	
TCAGACTGTT	AATACAGATT	TCTCTCCATA	TCTGATTICA	GATAACTTAG	AACAGCCTAT	2940	
GGGAAGTAGT	CATGCATCTC	AGGTTTGTTC	TGAGACACCT	GATGACCTGT	TAGATGATGG	3000	
TGAAATAAAG	GAAGATACTA	GTTTTGCTGA	AAATGACATT	AAGGAAAGTT	CTGCTGTTTT	3060	
TAGCAAAAGC	GTCCAGAAAG	GAGAGCTTAG	CAGGAGTCCT	AGCCCTTTCA	CCCATACACA	3120	
TTTGGCTCAG	GGTTACCGAA	GAGGGGCCAA	GAAATTAGAG	TCCTCAGAAAG	AGAACTTATC	3180	
TAGTGAGGAT	GAAGAGCTTC	CCTGCTTCCA	ACACTTGTTA	TTTGGTAAAG	TAAACAATAT	3240	
ACCTTCTCAG	TCTACTAGGC	ATAGCACCGT	TGCTACCGAG	TGTCTGTCTA	AGAACACAGA	3300	
GGAGAATTTA	TTATCATTGA	AGAATAGCTT	AAATGACTGC	AGTAACCAAG	TAATATTGGC	3360	
AAAGGCATCT	CAGGAACATC	ACCTTAGTGA	GGAAACAAAA	TGTTCTGTCTA	GCTTGTTTTTC	3420	
TTACAGTGC	AGTGAATTGG	AAGACTTGAC	TGCAAAATACA	AACACCCAGG	ATCCTTTCTT	3480	
GATTGGTTCT	TCCAAACAAA	TGAGGCATCA	GTCTGAAAAG	CAGGGAGTTG	GTCTGAGTGA	3540	
CAAGGAATTG	GTTTCAGATG	ATGAAGAAAG	AGGAACGGGC	TTGGAAGAAA	ATAATCAAGA	3600	
AGAGCAAAAG	ATGGATTCAA	ACTTAGGTAT	TGGAACCAAG	TTTTTGTGTT	TGCCCCAGTC	3660	
TATTTATAGA	AGTGAAGCTA	ATGTTTATGC	TTTTGGGGAG	CACATTTTAC	AAATTTCCAA	3720	
GTATAGTTAA	AGGAAC TGCT	TCTTAACTT	GAAACATGTT	CCTCCTAAGG	TGCTTTTCAT	3780	
AGAAAAAGT	CCTTCACACA	GCTAGGACGT	CATCTTTGAC	TGAATGAGCT	TTAACATCCT	3840	
AATTACTGGT	GGACTTACTT	CTGGTTTTCAT	TTTATAAAGC	AAATCCCGGT	GTCCCAAAGC	3900	
AAGGAATTTA	ATCATTTTGT	GTGACATGAA	AGTAAATCCA	GTCTTGCCAA	TGAGAAGAAA	3960	
AAGACACAGC	AAGTTGCAGC	GTTTATAGTC	TGCTTTTACA	TCTGAACCTC	TGTTTTTGTT	4020	
ATTTAAGGTG	AAGCAGCATC	TGGGTGTGAG	AGTGAAACAA	GCGTCTCTGA	AGACTGCTCA	4080	
GGGCTATCCT	CTCAGAGTGA	CATTTTAAAC	ACTCAGGTAA	AAAGCGTGTG	TGTGTGTGCA	4140	
CATGCGTGTG	TGTGGTGTCC	TTTGCAATTCA	GTAGTATGTA	TCCACATTC	TTAGGTTTGC	4200	
TGACATCATC	TCITTGAATT	AATGGCACAA	TTGTTTGTGG	TTCATTGTC		4249	

5,747,282

115

116

-continued

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 710 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(i v) ANTI-SENSE: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: *Homo sapiens*

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:22:

NNGAATGTGA	ATCCTAAIAT	TTCNCNCCNA	CTTAAAAGAA	TACCACTCCA	ANGGCATCNC	60
AATACATCAA	TCAATTGGGG	AATTGGGATT	TTCCCTCNCI	AACATCANTG	GAATAAATTC	120
ATGGCATTAA	TTGCATGAAT	GTGGTTAGAT	TAAAAGGTGT	TCATGCTAGA	ACTTGTAGTT	180
CCATACTAGG	TGATTTCAAT	TCCTGTGCTA	AAATTAATTT	GTATGATATA	TTNTCATTTA	240
ATGGAAAGCT	TCTCAAAGTA	TTTCATTTTC	TTGGTACCAT	TTATCGTTTT	TGAAGCAGAG	300
GGATACCATG	CAACATAACC	TGATAAAGCT	CCAGCAGGAA	ATGGCTGAAC	TAGAAGCTGT	360
GTTAGAACAG	CATGGGAGCC	AGCCTTCTAA	CAGCTACCCT	TCCATCATAA	GTGACTCTTC	420
TGCCCTTGAG	GACCTGCGAA	ATCCAGAACA	AAGCACATCA	GAAAAAGGTG	TGTATTGTTG	480
GCCAAACACT	GATATCTTAA	GCAAAATTCT	TTCTTCCCC	TTTATCTCCT	TCTGAAGAGT	540
AAGGACCTAG	CTCCAACATT	TTATGATCCT	TGCTCAGCAC	ATGGGTAATT	ATGGAGCCTT	600
GGTTCCTTGC	CCTGCTCACA	ACTAATATAC	CAGTCAGAGG	GACCCAAGGC	AGTCATTTCAT	660
GTTGTCACTC	GAGATACCTA	CAACAAGTAG	ATGCTATGGG	GAGCCCATGG		710

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 473 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(i v) ANTI-SENSE: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: *Homo sapiens*

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:23:

CCATTGGTGC	TAGCATCTGT	CTGTTGCATT	GCTGTGTTTT	ATAAAATTCT	GCCTGATATA	60
CTTGTTAAAA	ACCAATTITG	GTATCATAGA	TTGATGCTTT	TGAAAAAAAT	CAGTATTCTA	120
ACCTGAATTA	TCATATCAG	AACAAAGCAG	TAAAGTAGAT	TTGTTTTCTC	ATTCCATTTA	180
AAGCAGTATT	AAC TTCACAG	AAAAGTAGTG	AATACCCTAT	AAGCCAGAAT	CCAGAAGGCC	240
TTTCTGCTGA	CAAGTTTGA	GTGCTGCGAG	ATAGTTCTAC	CAGTAAAAAT	AAAGAACCAG	300
GAGTGGAAAG	GTAAGAAACA	TCAATGTAAA	GATGCTGTGG	TATCTGACAT	CTTTATTTAT	360
ATTGAACTCT	GATTGTTAAT	TTTTTTCACC	ATACTTTCTC	CAGTTTTTTT	GCATACAGGC	420
ATTTATACAC	TTTTATTGCT	CTAGGATACT	TCIIITGTTT	AATCCTATAT	AGG	473

5,747,282

117

118

-continued

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 421 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(i v) ANTI-SENSE: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: *Homo sapiens*

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:24:

GGATAAGNTC	AAGAGATATT	TTGATAGGTG	ATGCAGTGAT	NAATTONGAA	AATTTNCTGC	60
CTGCTTTTAA	TCTTCCCCCG	TTCTTTCTTC	CTNCCTCCCT	CCCTTCCTNC	CTCCCGTCCT	120
TNCCTTTTCT	TTCCCTCCCT	TCCNCCTTCT	TTCCNICTNT	CTTTCCTTTC	TTTCCGTGCT	180
ACCTTTCTTT	CCTTCCTCCC	TTCTTTTCT	TTTCTTTCTT	TCCTTTCCTT	TTCTTTCCTT	240
TCCTTTCCTT	CCTTTCCTTC	TTGACAGAGT	CTTGTCTGT	CACTCAGGCT	GGAGTGCAGT	300
GGCGTGATCT	CGNCTCACTG	CAACCTCTGT	CTCCCAGGTT	CAAGCAATTT	TCCTGCCTCA	360
GCCTCCCGAG	TAGCTGAGAT	TACAGGCGCC	AGCCACCACA	CCCAGCTACT	GACCTGCTTT	420
T						421

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 997 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(i v) ANTI-SENSE: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: *Homo sapiens*

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:25:

AAAAAGCTGG	GAGATATGGT	GCCTCAGACC	AACCCCATGT	TATATGTCAA	CCCTGACATA	60
TTGGCAGGCA	ACATGAATCC	AGACTTCTAG	GCTGTCTATGC	GGGCTCTTTT	TTGCCAGTCA	120
TTTCTGATCT	CTCTGACATG	AGCTGTTTCA	TTTATGCTTT	GGCTGCCCAG	CAAGTATGAT	180
TTGTCTTTTC	ACAATTGGTG	GCGATGGTTT	TCTCCTTCCA	TTTATCTTTT	TAGGTCATCC	240
CCTTCTAAAT	GCCCATCATT	AGATGATAGG	TGGTACATGC	ACAGTTGCTC	TGGGAGTCTT	300
CAGAATAGAA	ACTACCCATC	TCAAGAGGAG	CTCATTAAGG	TTGTTGATGT	GGAGGAGCAA	360
CAGCTGGAAG	AGTCTGGGCC	ACACGATTTG	ACGGAACAT	CTTACTTGCC	AAGGCAAGAT	420
CTAGGTAATA	TTTCATCTGC	TGTATTGGAA	CAAACACT Y T	GATTTTACTC	TGAATCCTAC	480
ATAAAGATAT	TCTGGTTAAC	CAACTTTTAG	ATGTACTAGT	CTATCATGGA	CACTTTTGTT	540
ATACTTAATT	AAGCCCACTT	TAGAAAAATA	GCTCAAGTGT	TAATCAAGGT	TTACTTGAAA	600
ATTATTGAAA	CTGTTAATCC	ATCTATATTT	TAATTAATGG	TTTAACTAAT	GATTTTGAGG	660
ATGWWGGAGT	CKTGGTGTAC	TCTAMATGTA	TTATTTTCAGG	CCAGGCATAG	TGGCTCACGC	720
CTGGTAATCC	CAGTA Y Y CMR	GAGCCCGAGG	CAGGTGGAGC	CAGCTGAGGT	CAGGAGTTCA	780

5,747,282		
119		120
-continued		
<hr/>		
AGACCTGTCT TGGCCAACAT GGGNGAAACC CTGICTTCTT CTIAAAAAAN ACAAAAAAAA	840	
TTAACTGGGT TGIGCTTAGG TGNATGCCCC GNATCCTAGT TTTCTTNG GGTGAGGGA	900	
GGAGATCACN TTGGACCCCG GAGGGGNGGG TGGGGGNGAG CAGGNCAAAA CACNGACCCA	960	
GCTGGGGTGG AAGGGAAGCC CACTCNAAAA AANNITN	997	
(2) INFORMATION FOR SEQ ID NO:26:		
(i) SEQUENCE CHARACTERISTICS:		
(A) LENGTH: 639 base pairs		
(B) TYPE: nucleic acid		
(C) STRANDEDNESS: double		
(D) TOPOLOGY: linear		
(i i) MOLECULE TYPE: DNA (genomic)		
(i i i) HYPOTHETICAL: NO		
(i v) ANTI-SENSE: NO		
(v i) ORIGINAL SOURCE:		
(A) ORGANISM: Homo sapiens		
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:26:		
TTTTTAGGAA ACAAGCTACT TTGGATTTC ACCAACACCT GTATTCAITGT ACCCATTTTT	60	
CTCTTAACCT AACTTTATTG GTCTTTTAA TTCITTAACAG AGACCAGAAC TTGTAAATTC	120	
AACATTCATC GTTGTGTAAT TAAACTTCT CCCATTCTT TCAGAGGGAA CCCCTTACCT	180	
GGAATCTGGA ATCAGCCTCT TCTCTGATGA CCTGAATCT GATCCTTCTG AAGACAGAGC	240	
CCCAGAGTCA GCTCGTGTTG GCAACATACC ATCTTCAACC TCTGCATTGA AAGTTCCCCA	300	
ATTGAAAGTT GCAGAATCTG CCCAGAGTCC AGCTGCTGCT CATACTACTG ATACTGCTGG	360	
GTATAATGCA ATGGAAGAAA GTGTGAGCAG GGAGAAGCCA GAATTGACAG CTTCAACAGA	420	
AAGGGTCAAC AAAAGAATGT CCATGGTGGT GTCTGGCCTG ACCCCAGAAG AATTTGTGAG	480	
TGTATCCATA TGTATCTCCC TAATGACTAA GACTTAACAA CATTCTGGAA AGAGTTTTAT	540	
GTAGGTATTG TCAATTAATA ACCTAGAGGA AGAAATCTAG AAAACAATCA CAGTTCTGTG	600	
TAATTTAATT TCGATTACTA ATTTCTGAAA ATTTAGAA Y	639	
(2) INFORMATION FOR SEQ ID NO:27:		
(i) SEQUENCE CHARACTERISTICS:		
(A) LENGTH: 922 base pairs		
(B) TYPE: nucleic acid		
(C) STRANDEDNESS: double		
(D) TOPOLOGY: linear		
(i i) MOLECULE TYPE: DNA (genomic)		
(i i i) HYPOTHETICAL: NO		
(i v) ANTI-SENSE: NO		
(v i) ORIGINAL SOURCE:		
(A) ORGANISM: Homo sapiens		
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:27:		
NCCCNCCCC CNAATCTGAA ATGGGGGTAA CCCCCCCCCA ACCGANACNT GGGTNGCNTA	60	
GAGANTTTAA TGGCCNTTC TGAAGNACAN AAGCTTAAGC CAGGNGACGT GGANCNATGN	120	
GTTGTITNTT GTTGGTTAC CTCCAGCCTG GGTGACAGAG CAAGACTCTG TCTAAAAAAA	180	
AAAAAAAAAA AAATCGACTT TAAATAGTTC CAGGACACGT GTAGAACGTG CAGGATTGCT	240	
ACGTAGGTAA ACATATGCCA TGGTGGGATA ACTAGTATTC TGAGCTGTGT GCTAGAGGTA	300	

5,747,282

121

122

-continued

ACTCATGATA	ATGGAATATT	TGATTTAATT	TCAGATGCTC	GTGTACAAGT	TTGCCAGAAA	360
ACACCCACATC	ACTTTAACTA	ATCTAATTAC	TGAAGAGACT	ACTCATGTTG	TTATGAAAAC	420
AGGTATACCA	AGAACCCTTTA	CAGAATACCT	TGCATCTGCT	GCATAAAACC	ACATGAGGCG	480
AGGCACGGTG	GCGCATGCCT	GTAATCGCAG	CACCTTGGGA	GGCCGAGGCG	GGCAGATCAC	540
GAGATTAGGA	GATCGAGACC	ATCCTGGCCA	GCATGGTGAA	ACCCCGTCTC	TACTANNAAA	600
TGGNAAAATT	ANCTGGGTGT	GGTCGCGTGC	NCCTGTAGTC	CCAGCTACTC	GTGAGGCTGA	660
GGCAGGAGAA	TCACTTGAAC	CGGGGAAATG	GAGGTTTCAG	TGAGCAGAGA	TCATNCCCCT	720
NCATTCCAGC	CTGGCGACAG	AGCAAGGCTC	CGTCNCCNAA	AAAATAAAAA	AAAACGTGAA	780
CAATAAAGAA	TATTTGTTGA	GCATAGCATG	GATGATAGTC	TTCTAATAGT	CAATCAATTA	840
CTTTATGAAA	GACAAATAAT	AGTTTTGCTG	CTTCCTTACC	TCCTTTTGTT	TTGGGTTAAG	900
ATTTGGAGTG	TGGGCCAGGC	AC				922

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 867 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(i v) ANTI-SENSE: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: *Homo sapiens*

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:28:

GATCTATAGC	TAGCCTTGGC	GTCTAGAAGA	TGGGTGTTGA	GAAGAGGGAG	TGGAAAGATA	60
TTTCCTCTGG	TCTTAACTTC	ATATCAGCCT	CCCCTAGACT	TCCAAATATC	CATACCTGCT	120
GGTTATAATT	AGTGGTGTTC	TCAGCCTCTG	ATTCTGTAC	CAGGGGTTTT	AGAATCATAA	180
ATCCAGATTG	ATCTTGGGAG	TGAAAAAAC	TGAGGCTCTT	TAGCTTCTTA	GGACAGCACT	240
TCCTGATTTT	GTTTTCAACT	TCTAATCCTT	TGAAGTGTTC	TCATTCTGCA	GATGCTGAGT	300
TTGTGTGTGA	ACGGACACTG	AAATATTTTC	TAGGAATTGC	GGGAGGAAAA	TGGGTAGTTA	360
GCTATTTCTG	TAAGTATAAT	ACTATTTCTC	CCCTCCTCCC	TTTAACACCT	CAGAATTGCA	420
TTTTTACACC	TAACATTTAA	CACCTAAGGT	TTTTGCTGAT	GCTGAGTCTG	AGTTACCAAA	480
AGGTCTTTAA	ATTGTAATAC	TAAACTACTT	TTATCTTTAA	TATCACTTTG	TTCAAGATAA	540
GCTGGTGATG	CTGGGAAAAAT	GGGTCTCTTT	TATAACTAAT	AGGACCTAAT	CTGCTCCTAG	600
CAATGTTAGC	ATATGAGCTA	GGGATTTATT	TAATAGTCGG	CAGGAATCCA	TGTGCARCAG	660
NCAAACTTAT	AATGTTTAAA	TTAAACATCA	ACTCTGTCTC	CAGAAGGAAA	CTGCTGCTAC	720
AAGCCTTATT	AAAGGGCTGT	GGCTTTAGAG	GGAAGGACCT	CTCCTCTGTC	ATTCTTCCTG	780
TGCTCTTTTG	TGAATCCTG	ACCTCTCTAT	CTCCGTGAAA	AGAGCACGTT	CTTCTGCTGT	840
ATGTAACCTG	TCTTTTCTAT	GATCTCT				867

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 561 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double

5,747,282

123

124

-continued

(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(i v) ANTI-SENSE: NO

(v i) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:29:

NAAAAACGGG GNNGGGANTG GGCCTTAAAN CCAAAGGGCN AACTCCCCAA CCATTNAAAA 60

ANTGACNGGG GATTATTAAA ANCGGCGGGA AACATTTAC NGCCCAACTA ATATTGTTAA 120

ATTAAAACCA CCACCNCTGC NCCAAGGAGG GAAACTGCTG CTACAAGCCT TATTAAAGGG 180

CTGTGGCTTT AGAGGGAAAG ACCTCTCCTC TGTCACTCTT CCTGTGCTCT TTTGTGAATC 240

GCTGACCCTT CTATGTCCTG GAAAAGAGCA CGTTCCTCTG CTGTATGTAA CCTGTCTTTT 300

CTATGATCTC TTTAGGGGTG ACCCAGTCTA TTAAAGAAAAG AAAAATGCTG AATGAGGTAA 360

GTACTTGATG TTACAACTA ACCAGAGATA TTCATTCACT CATATAGTTA AAAATGTATT 420

TGCTTCCTTC CATCAATGCA CCACCTTCCT TAACAATGCA CAAATTTTCC ATGATAATGA 480

GGATCATCAA GAATTATGCA GGCCTGCACT GTGGCTCATA CCTATAATCC CAGCGCTTTG 540

GGAGGCTGAG GCGCTTGGAT C 561

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 567 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(i v) ANTI-SENSE: NO

(v i) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:30:

AATTTTTTGT ATTTTATGTA GAGATGAGGT TCACCATGTT GGTCTAGATC TGGTGTGCGAA 60

CGTCCTGACC TCAAGTGATC TGCCAGCCTC AGTCTCCCAA AGTGCTAGGA TTACAGGGGT 120

GAGCCACTGC GCCTGGCCTG AATGCCTAAA ATATGACGTG TCTGCTCCAC TTCCATTGAA 180

GGAAGCTTCT CTTTCTCTTA TCCTGATGGG TTGTGTTTGG TTTCTTTTCA CATGATTTTG 240

AAGTCAGAGG AGATGTGGTC AATGGAAGAA ACCACCAAGG TCCAAAGCGA GCAAGAGAAT 300

CCCAGGACAG AAAGGTAAAG CTCCTCCCT CAAGTTGACA AAAATCTCAC CCCACCACTC 360

TGTATTCCAC TCCCCTTTGC AGAGATGGGC CGCTTCATTT TGTAAGACTT ATTACATACA 420

TACACAGTGC TAGATACTTT CACACAGGTT CTTTTTTCAC TCTTCCATCC CAACCACATA 480

AATAAGTATT GTCTCTACTT TATGAATGAT AAAACTAAGA GATTTAGAGA GGCTGTGTAA 540

TTTGGATTCC CGTCTCGGGT TCAGATC 567

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 633 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

5,747,282

125

126

-continued

(D) TOPOLOGY: linear	
(i i) MOLECULE TYPE: DNA (genomic)	
(i i i) HYPOTHETICAL: NO	
(i v) ANTI-SENSE: NO	
(v i) ORIGINAL SOURCE:	
(A) ORGANISM: Homo sapiens	
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:31:	
TTGGCCTGAT TGGTGACAAA AGTGAGATGC TCAGTCCTTG AATGACAAAAG AATGCCTGTA	60
GAGTTGCAGG TCCAATACTA TATGCACTTC AAGAAGATCT TCTGAAATCT AGTAGTGTTT	120
TGGACATTGG ACTGCTTTGTC CCTGGGAAAGT AGCAGCAGAA ATGATCGGTG GTGAACAGAA	180
GAAGAAAGAAA AGCTCTTCCT TTTTGAAAGT CTGTTTTTTG AATAAAAGCC AATATTCITT	240
TATAACTAGA TTTTCCTTCT CTCCATTCCC CTGTCCCTCT CTCTTCTCTCT CTTCTTCCAG	300
ATCTTCAGGG GGCTAGAAAT CTGTTGCTAT GGGCCCTTCA CCAACATGCC CACAGGTAAG	360
AGCCTGGGAG AACCCAGAG TTCCAGCACC AGCCTTTGTC TTACATAGTG GAGTATTATA	420
AGCAAGGTCC CACGATGGGG GTTCCTCAGA TTGCTGAAAT GTTCTAGAGG CTATTCTATT	480
TCTCTACCAC TCTCCAAACA AAACAGCACC TAAATGTTAT CCTATGGCAA AAAAAAATA	540
TACCTTGICC CCCTTCTCAA GAGCATGAAAG GTGGTTAATA GTTAGGATTG AGTATGTTAT	600
GTGTTTCAAT GGCCTTGAGC TGCTGTTAGT GCC	633
(2) INFORMATION FOR SEQ ID NO:32:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 470 base pairs	
(B) TYPE: nucleic acid	
(C) STRANDEDNESS: double	
(D) TOPOLOGY: linear	
(i i) MOLECULE TYPE: DNA (genomic)	
(i i i) HYPOTHETICAL: NO	
(i v) ANTI-SENSE: NO	
(v i) ORIGINAL SOURCE:	
(A) ORGANISM: Homo sapiens	
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:32:	
TTTGAGAGAC TATCAAACCT TATACCAAGT GGCCTTATGG AGACTGATAA CCAGAGTACA	60
TGGCATATCA GTGGCAAATT GACTTAAAT CCATACCCCT ACTATTTTAA GACCATTGTC	120
CTTTGGAGCA GAGAGACAGA CTCTCCCAT TGGAGGCTCT GCTATAAGCC TTCATCCGGA	180
GAGTGTAGGG TAGAGGGCCT GGGTTAAGTA TGCAGATTAC TGCAGTGATT TTACATGTAA	240
ATGTCCATTT TAGATCAACT GGAATGGATG GTACAGCTGT GTGGTGCTTC TGTGGTGAAAG	300
GAGCTTTTCA CATTACCCCT TGGCACAGTA AGTATTGGGT GCCCTGTGAG TGTGGGAGGA	360
CACAATATTC TCTCCTGTGA GCAAGACTGG CACCTGTGAG TCCCTATGGA TGCCCCACT	420
GTAGCCTCAG AAGTCTTCTC TGCCACATA CCTGTGCCAA AAGACTCCAT	470
(2) INFORMATION FOR SEQ ID NO:33:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 517 base pairs	
(B) TYPE: nucleic acid	
(C) STRANDEDNESS: double	
(D) TOPOLOGY: linear	

5,747,282

127

128

-continued

```
( i i ) MOLECULE TYPE: DNA (genomic)

( i i i ) HYPOTHETICAL: NO

( i v ) ANTI-SENSE: NO

( v i ) ORIGINAL SOURCE:
      ( A ) ORGANISM: Homo sapiens

( x i ) SEQUENCE DESCRIPTION: SEQ ID NO:33:

GGTGGTACGT  GICTGTAGTT  CCAGCTACTT  GGGAGGCTGA  GATGGAAGGA  TTGCTTGAGC          60
CCAGGAGGCA  GAGGTGGNAN  NTTACGCTGA  GATCACACCA  CTGCACTCCA  GCCTGGGTGA          120
CAGAGCAAGA  CCCTGTCTCA  AAAACAAACA  AAAAAAATGA  TGAAGTGACA  GTTCCAGTAG          180
TCCTACTTTG  ACACTTTGAA  TGCTCTTTCC  TTCCTGGGGA  TCCAGGGTGT  CCACCCAATT          240
GTGGTTGTGC  AGCCAGATGC  CTGGACAGAG  GACAATGGCT  TCCATGGTAA  GGTGCCTCGC          300
ATGTACCTGT  GCTATTAGTG  GGGTCCTTGT  GCATGGGTTT  GGTTTATCAC  TCATTACCTG          360
GTGCTTGAGT  AGCACAGTTC  TTGGCACATT  TTTAAATATT  TGTTGAATGA  ATGGCTAAAA          420
TGTCTTTTTT  ATGTTTTTAT  TGTATTATTT  TTTATATTGT  AAAAGTAATA  CATGAACGTG          480
TTCCATGGGG  TGGGAGTAA  ATATGAATGT  TCAACAC          510
```

(2) INFORMATION FOR SEO ID NO:34:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 434 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL; NO

(i v) ANTI-SENSE: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: *Homo sapiens*

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:34:

CAGTAATCCT	NAGAACTCAT	ACGACCGGGC	CCCTGGAGTC	GNTGNTTNGA	GCCTAGTCCN	60
GGAGAATGAA	TTGACACTAA	TCTCTGCTTG	TGTTCTCTGT	CTCCAGCAAT	TGGGCAGATG	120
TGTGAGGCAC	CTGTGGTGAC	CCGAGAGTGG	GTGTGGACA	GTGTAGCACT	CTACCAGTGC	180
CAGGAGCTGG	ACACCTACCT	GATACCCAG	ATCCCCACA	GCCACTACTG	ACTGCAGCCA	240
GCCACAGGTA	CAGAGCCACA	GGACCCCAAG	AATGAGCTTA	CAAAGTGGCC	TTTCCAGGCC	300
CTGGGAGCTC	CTCTCACTCT	TCAGTCCTTC	TACTGTCTTG	GCTACTAAAT	ATTTTATGTA	360
CATCAGCCTG	AAAAGGACTT	CTGGCTATGC	AAGGGTCCCT	TAAAGATTTT	CTGCTTGAAG	420
TCTCCCTTGG	AAAT					434

(2) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

5,747,282

129

130

-continued

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:35:
GATAAATTAA AACTGCGACT GCGCGGCGTG 3 0

(2) INFORMATION FOR SEQ ID NO:36:
(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 30 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
(i i) MOLECULE TYPE: DNA (genomic)
(i i i) HYPOTHETICAL: NO
(v i) ORIGINAL SOURCE:
 (A) ORGANISM: Homo sapiens
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:36:
GTAGTAGAGT CCCGGGAAAG GGACAGGGGG 3 0

(2) INFORMATION FOR SEQ ID NO:37:
(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 30 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
(i i) MOLECULE TYPE: DNA (genomic)
(i i i) HYPOTHETICAL: NO
(v i) ORIGINAL SOURCE:
 (A) ORGANISM: Homo sapiens
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:37:
ATATATATAT GTTTTCTAA TGTGTTAAAG 3 0

(2) INFORMATION FOR SEQ ID NO:38:
(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 30 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
(i i) MOLECULE TYPE: DNA (genomic)
(i i i) HYPOTHETICAL: NO
(v i) ORIGINAL SOURCE:
 (A) ORGANISM: Homo sapiens
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:38:
GTAAGTCAGC ACAAGAGTGT ATTAATTGG 3 0

(2) INFORMATION FOR SEQ ID NO:39:
(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 30 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
(i i) MOLECULE TYPE: DNA (genomic)
(i i i) HYPOTHETICAL: NO
(v i) ORIGINAL SOURCE:
 (A) ORGANISM: Homo sapiens

5,747,282

131

132

-continued

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:39:
TTTCTTTTTC TCCCCCCCCT ACCCTGCTAG 3 0

(2) INFORMATION FOR SEQ ID NO:40:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(i i) MOLECULE TYPE: DNA (genomic)
(i i i) HYPOTHETICAL: NO
(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:40:
GTAAGTTTGA ATGTGTTATG TGGCTCCATT 3 0

(2) INFORMATION FOR SEQ ID NO:41:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(i i) MOLECULE TYPE: DNA (genomic)
(i i i) HYPOTHETICAL: NO
(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:41:
AGCTACTTTT TTTTTTTTTT TTTGAGACAG 3 0

(2) INFORMATION FOR SEQ ID NO:42:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(i i) MOLECULE TYPE: DNA (genomic)
(i i i) HYPOTHETICAL: NO
(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:42:
GTAAGTGCAC ACCACCATAT CCAGCTAAAT 3 0

(2) INFORMATION FOR SEQ ID NO:43:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(i i) MOLECULE TYPE: DNA (genomic)
(i i i) HYPOTHETICAL: NO
(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

5,747,282

133

134

-continued

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:43:
AATTGTTCTT TCTTTCITTA TAATTTATAG 3 0

(2) INFORMATION FOR SEQ ID NO:44:
(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 30 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
(i i) MOLECULE TYPE: DNA (genomic)
(i i i) HYPOTHETICAL: NO
(v i) ORIGINAL SOURCE:
 (A) ORGANISM: Homo sapiens
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:44:
GTATATAATT TGGTAATGAT GCTAGGTTGG 3 0

(2) INFORMATION FOR SEQ ID NO:45:
(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 30 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
(i i) MOLECULE TYPE: DNA (genomic)
(i i i) HYPOTHETICAL: NO
(v i) ORIGINAL SOURCE:
 (A) ORGANISM: Homo sapiens
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:45:
GAGTGTGTTT CTCAAACAAT TTAATTTCAg 3 0

(2) INFORMATION FOR SEQ ID NO:46:
(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 30 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
(i i) MOLECULE TYPE: DNA (genomic)
(i i i) HYPOTHETICAL: NO
(v i) ORIGINAL SOURCE:
 (A) ORGANISM: Homo sapiens
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:46:
GTAAGTGTTG AATATCCCAA GAATGACACT 3 0

(2) INFORMATION FOR SEQ ID NO:47:
(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 30 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
(i i) MOLECULE TYPE: DNA (genomic)
(i i i) HYPOTHETICAL: NO
(v i) ORIGINAL SOURCE:
 (A) ORGANISM: Homo sapiens

5,747,282

135

136

-continued

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:47:
AAACATAATG TTTTCCTTG TATTTTACAG 3 0

(2) INFORMATION FOR SEQ ID NO:48:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:48:
GTAAACCAT TTGTTTCTT CTCTTCTTC 3 0

(2) INFORMATION FOR SEQ ID NO:49:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:49:
TGCTTGACTG TTCTTTACCA TACTGTTTAG 3 0

(2) INFORMATION FOR SEQ ID NO:50:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:50:
GTAAGGGTCT CAGGTTTTTT AAGTATTTAA 3 0

(2) INFORMATION FOR SEQ ID NO:51:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

5,747,282

137

138

-continued

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:51:

TGATTTAATTI TTGGGGGGA AATTTTTTAG 30

(2) INFORMATION FOR SEQ ID NO:52:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: *Homo sapiens*

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:52:

G T G A G T C A A A G A G A A C C T T T G T C T A T G A A G 30

(2) INFORMATION FOR SEQ ID NO:53:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) **HYPOTHETICAL: NO**

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:53:

TCCTATTAGG ACTCTGTCCT TTCCCTATAG 30

(2) INFORMATION FOR SEQ ID NO:54:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) **HYPOTHETICAL: NO**

(v i) ORIGINAL SOURCE:
(A) ORGANISM: *Homo sapiens*

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:54:

GTAATGGCAA AGTTTGCCAA CTTAACAGGC 30

(2) INFORMATION FOR SEQ ID NO:55:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

5,747,282

139

140

-continued

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:55:
GAGTACCTTG TTATTTTGT ATATTTTCAG 3 0

(2) INFORMATION FOR SEQ ID NO:56:
(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 30 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
(i i) MOLECULE TYPE: DNA (genomic)
(i i i) HYPOTHETICAL: NO
(v i) ORIGINAL SOURCE:
 (A) ORGANISM: Homo sapiens
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:56:
GTATTGGAAC CAGGTTTTTG TGTTCGCC 3 0

(2) INFORMATION FOR SEQ ID NO:57:
(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 30 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
(i i) MOLECULE TYPE: DNA (genomic)
(i i i) HYPOTHETICAL: NO
(v i) ORIGINAL SOURCE:
 (A) ORGANISM: Homo sapiens
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:57:
ACATCTGAAC CTCTGTTTT GTATTTAAG 3 0

(2) INFORMATION FOR SEQ ID NO:58:
(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 30 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
(i i) MOLECULE TYPE: DNA (genomic)
(i i i) HYPOTHETICAL: NO
(v i) ORIGINAL SOURCE:
 (A) ORGANISM: Homo sapiens
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:58:
AGGTAAAAAG CGTGTGTGTG TGTGCACATG 3 0

(2) INFORMATION FOR SEQ ID NO:59:
(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 30 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
(i i) MOLECULE TYPE: DNA (genomic)
(i i i) HYPOTHETICAL: NO
(v i) ORIGINAL SOURCE:
 (A) ORGANISM: Homo sapiens

5,747,282

141

142

-continued

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:59:
CATTTCCTTG GTACCATTTA TCGTTTTTGA 3 0

(2) INFORMATION FOR SEQ ID NO:60:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:60:
GTGTGTATTG TTGGCCAAAC ACTGATATCT 3 0

(2) INFORMATION FOR SEQ ID NO:61:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:61:
AGTAGATTTG TTTTCTCATT CCAITTTAAAG 3 0

(2) INFORMATION FOR SEQ ID NO:62:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:62:
GTAAGAAACA TCAATGTAAA GATGCTGTGG 3 0

(2) INFORMATION FOR SEQ ID NO:63:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

5,747,282

143

144

-continued

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:63:

ATGGTTTTCT CCTTCCATTT ATCTTTCTAG

30

(2) INFORMATION FOR SEO ID NO:64:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: *Homo sapiens*

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:64:

GTAATATTTC ATCTGCTGTA TTGGAACAAA

30

(2) INFORMATION FOR SEQ ID NO:65:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: *Homo sapiens*

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:65:

TGTA AATTAA ACTTCTCCCA TTCCTTTCAG

30

(2) INFORMATION FOR SEQ ID NO:66:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:66:

GTGAGTGTAT CCATATGTAT CTCCTAATG

30

(2) INFORMATION FOR SEQ ID NO:67:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

5,747,282

145

146

-continued

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:67:
ATGATAATGG AATATTTGAT TTAATTTTCAG 3 0

(2) INFORMATION FOR SEQ ID NO:68:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:68:
GTATACCAAG AACCTTTACA GAATACCTTG 3 0

(2) INFORMATION FOR SEQ ID NO:69:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:69:
CTAATCCTTT GAGTGTTTT CATTCTGCAG 3 0

(2) INFORMATION FOR SEQ ID NO:70:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:70:
GTAAGTATAA TACTATTTCT CCCCTCCTCC 3 0

(2) INFORMATION FOR SEQ ID NO:71:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

5,747,282

147

148

-continued

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:71:
TGTAACCTGT CTTTCTATG ATCTCTTTAG 3 0

(2) INFORMATION FOR SEQ ID NO:72:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:72:
GTAAGTACTT GATGTTACAA ACTAACCAGA 3 0

(2) INFORMATION FOR SEQ ID NO:73:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:73:
TCCTGATGGG TTGTGTTTGG TTTCTTTCAG 3 0

(2) INFORMATION FOR SEQ ID NO:74:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:74:
GTAAAGCTCC CTCCTCAAG TTGACAAAAA 3 0

(2) INFORMATION FOR SEQ ID NO:75:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

5,747,282

149

150

-continued

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:75:
CTGTCCCTCT CTCCTCCTCT CTCCTTCCAG 3 0

(2) INFORMATION FOR SEQ ID NO:76:
(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 30 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
(i i) MOLECULE TYPE: DNA (genomic)
(i i i) HYPOTHETICAL: NO
(v i) ORIGINAL SOURCE:
 (A) ORGANISM: Homo sapiens
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:76:
GTAAGAGCCT GGGAGAACCC CAGAGTTCCA 3 0

(2) INFORMATION FOR SEQ ID NO:77:
(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 30 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
(i i) MOLECULE TYPE: DNA (genomic)
(i i i) HYPOTHETICAL: NO
(v i) ORIGINAL SOURCE:
 (A) ORGANISM: Homo sapiens
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:77:
AGTGATTTTA CATGTAAATG TCCATTTTAG 3 0

(2) INFORMATION FOR SEQ ID NO:78:
(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 30 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
(i i) MOLECULE TYPE: DNA (genomic)
(i i i) HYPOTHETICAL: NO
(v i) ORIGINAL SOURCE:
 (A) ORGANISM: Homo sapiens
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:78:
GTAAGTATTG GGTGCCCTGT CAGTGTGGGA 3 0

(2) INFORMATION FOR SEQ ID NO:79:
(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 30 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
(i i) MOLECULE TYPE: DNA (genomic)
(i i i) HYPOTHETICAL: NO
(v i) ORIGINAL SOURCE:
 (A) ORGANISM: Homo sapiens

5,747,282

151

152

-continued

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:79:
TTGAATGCTC TTTCCTTTCCT GGGGATCCAG 3 0

(2) INFORMATION FOR SEQ ID NO:80:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(i i) MOLECULE TYPE: DNA (genomic)
(i i i) HYPOTHETICAL: NO
(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:80:
GTAAGGTGCC TCGCATGTAC CTGTGCTATT 3 0

(2) INFORMATION FOR SEQ ID NO:81:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(i i) MOLECULE TYPE: DNA (genomic)
(i i i) HYPOTHETICAL: NO
(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:81:
CTAATCTCTG CTGTGTTTCT CTGTCTCCAG 3 0

(2) INFORMATION FOR SEQ ID NO:82:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 42 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear
(i i) MOLECULE TYPE: peptide
(i i i) HYPOTHETICAL: NO
(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:82:
Cys Pro Ile Cys Leu Glu Leu Ile Lys Glu Pro Val Ser Thr Lys Cys
1 5 10 15
Asp His Ile Phe Cys Lys Phe Cys Met Leu Lys Leu Leu Asn Gln Lys
20 25 30
Lys Gly Pro Ser Gln Cys Pro Leu Cys Lys
35 40

(2) INFORMATION FOR SEQ ID NO:83:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 45 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

5,747,282

153

154

-continued

(i i) MOLECULE TYPE: peptide

(i i i) HYPOTHETICAL: NO

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:83:

Cys	Pro	Ile	Cys	Leu	Glu	Leu	Leu	Lys	Glu	Pro	Val	Ser	Ala	Asp	Cys
1				5					10					15	
Asn	His	Ser	Phe	Cys	Arg	Ala	Cys	Ile	Thr	Leu	Asn	Tyr	Glu	Ser	Asn
			20					25					30		
Arg	Asn	Thr	Asp	Gly	Lys	Gly	Asn	Cys	Pro	Val	Cys	Arg			
		35					40					45			

(2) INFORMATION FOR SEQ ID NO:84:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 41 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS:

(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: peptide

(i i i) HYPOTHETICAL: NO

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:84:

Cys	Pro	Ile	Cys	Leu	Asp	Met	Leu	Lys	Asn	Thr	Met	Thr	Thr	Lys	Glu
1				5					10					15	
Cys	Leu	His	Arg	Phe	Cys	Ser	Asp	Cys	Ile	Val	Thr	Ala	Leu	Arg	Ser
			20					25					30		
Gly	Asn	Lys	Glu	Cys	Pro	Thr	Cys	Arg							
		35					40								

(2) INFORMATION FOR SEQ ID NO:85:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 42 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS:

(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: peptide

(i i i) HYPOTHETICAL: NO

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:85:

Cys	Pro	Val	Cys	Leu	Gln	Tyr	Phe	Ala	Glu	Pro	Met	Met	Leu	Asp	Cys
1				5					10					15	
Gly	His	Asn	Ile	Cys	Cys	Ala	Cys	Leu	Ala	Arg	Cys	Trp	Gly	Thr	Ala
			20					25					30		
Cys	Thr	Asn	Val	Ser	Cys	Pro	Gln	Cys	Arg						
		35					40								

What is claimed is:

1. An isolated DNA coding for a BRCA1 polypeptide, said polypeptide having the amino acid sequence set forth in SEQ ID NO:2.

2. The isolated DNA of claim 1, wherein said DNA has the nucleotide sequence set forth in SEQ ID NO:1.

3. The isolated DNA of claim 1 which contains BRCA1 regulatory sequences.

4. The isolated DNA of claim 2 which contains BRCA1 regulatory sequences.

5. An isolated DNA having at least 15 nucleotides of the DNA of claim 1.

55

6. An isolated DNA having at least 15 nucleotides of the DNA of claim 2.

7. An isolated DNA selected from the group consisting of:

(a) a DNA having the nucleotide sequence set forth in SEQ ID NO:1 having T at nucleotide position 4056;

(b) a DNA having the nucleotide sequence set forth in SEQ ID NO:1 having an extra C at nucleotide position 5385;

(c) a DNA having the nucleotide sequence set forth in SEQ ID NO: 1 having G at nucleotide position 5443;

and, (d) a DNA having the nucleotide sequence set forth in SEQ ID NO:1 having 11 base pairs at nucleotide positions 189-199 deleted.

A000241

5,747,282

155

8. A replicative cloning vector which comprises the isolated DNA of claim 1 or parts thereof and a replicon operative in a host cell.

9. A replicative cloning vector which comprises the isolated DNA of claim 2 or parts thereof and a replicon operative in a host cell.

10. An expression system which comprises the isolated DNA of claim 1 or parts thereof operably linked to suitable control sequences.

11. An expression system which comprises the isolated DNA of claim 2 or parts thereof operably linked to suitable control sequences.

12. Host cells transformed with the expression system of claim 10.

13. Host cells transformed with the expression system of claim 11.

14. A method of producing BRCA1 polypeptide which comprises culturing the cells of claim 12 under conditions effective for the production of said BRCA1 polypeptide and harvesting the BRCA1 polypeptide.

15. A method of producing BRCA1 polypeptide which comprises culturing the cells of claim 13 under conditions effective for the production of said BRCA1 polypeptide and harvesting the BRCA1 polypeptide.

16. A pair of single-stranded DNA primers for determination of a nucleotide sequence of a BRCA1 gene by a polymerase chain reaction, the sequence of said primers being derived from human chromosome 17q, wherein the use of said primers in a polymerase chain reaction results in the synthesis of DNA having all or part of the sequence of the BRCA1 gene.

156

17. The pair of primers of claim 16 wherein said BRCA1 gene has the nucleotide sequence set forth in SEQ ID NO:1.

18. A kit for detecting mutations in the BRCA1 gene resulting in a susceptibility to breast and ovarian cancers comprising at least one oligonucleotide primer specific for a BRCA1 gene mutation and instructions relating to detecting mutations in the BRCA1 gene.

19. A kit for detecting mutations in the BRCA1 gene resulting in a susceptibility to breast and ovarian cancers comprising at least one allele-specific oligonucleotide probe for a BRCA1 gene mutation and instructions relating to detecting mutations in the BRCA1 gene.

20. A method for screening potential cancer therapeutics which comprises: growing a transformed eukaryotic host cell containing an altered BRCA1 gene causing cancer in the presence of a compound suspected of being a cancer therapeutic, growing said transformed eukaryotic host cell in the absence of said compound, determining the rate of growth of said host cell in the presence of said compound and the rate of growth of said host cell in the absence of said compound and comparing the growth rate of said host cells, wherein a slower rate of growth of said host cell in the presence of said compound is indicative of a cancer therapeutic.

* * * * *

UNITED STATES PATENT AND TRADEMARK OFFICE
CERTIFICATE OF CORRECTION

PATENT NO. : 5,747,282
 DATED : May 5, 1998
 INVENTOR(S) : Mark H. Skolnick et al.

Page 1 of 6

It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

Title page,

Please add the following references:

--

U.S. PATENT DOCUMENTS

5,236,844 08/1993 Basset et al.

FOREIGN PATENT DOCUMENTS

0 518 650	12/1992	European Patent Office
91/09964	07/1991	WO
92/00311	01/1992	WO
92/11874	07/1992	WO
94/00764	01/1994	WO
95/19369	07/1995	WO

PUBLICATIONS

Castilla, L.H. et al. (1994). "Mutations in the BRCA 1 gene in families with early-onset breast and ovarian cancer," *Nature Genetics* 8:87-391.

Friedman, L.S. et al. (1994). "Confirmation of BRCA 1 by analysis of germline mutations linked to breast and ovarian cancer in ten families," *Nature Genetics* 8:399-404.

Goldgar, D.E. et al. (1994). "A Large Kindred With 17q-Linked Breast and Ovarian Cancer: Genetic, Phenotypic, and Genealogical Analysis," *J. Natl. Cancer Institute* 86:200-209.

Liang P. et al. (1992). "Differential Display and Cloning of Messenger RNAs from Human Breast Cancer versus Mammary Epithelial Cells," *Cancer Research* 52:6966-6968.

Neuhausen, S.L. et al. (1994). "A P1-based physical map of the region from D17S776 to D17378 containing the breast cancer susceptibility gene BRCA1," *Hum. Mol. Gen.* 3:1919-1926.

Sato, T. et al. (1992). "The Human Prohibitin Gene Located on Chromosome 17q21 is Mutated in Sporadic Breast Cancer," *Cancer Research* 52:1543-1646.

UNITED STATES PATENT AND TRADEMARK OFFICE
CERTIFICATE OF CORRECTION

PATENT NO. : 5,747,282
DATED : May 5, 1998
INVENTOR(S) : Mark. H. Skolnick et al.

Page 2 of 6

It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

Stampfer, M.R. et al. (1993). "Culture Systems for Study of Human Mammary Epithelial Cell Proliferation, Differentiation and Transformation," *Cancer Surveys* 18:7-34.

Wooster, R. et al. (1994). "Localization of a Breast Cancer Susceptibility Gene, BRCA2, to Chromosome 13q12-13," *Science* 265:2088-2090.

Albertsen, H., et al. (1994). "Genetic Mapping of the BRCA1 Region on Chromosome 17q21," *Am. J. Hum. Genet.* 54:516-525.

Anderson, D.E. (1972). "A Genetic Study of Human Breast Cancer," *J. Natl. Cancer Inst.* 48:1029-1034.

Arason, A., et al. (1993). "Linkage Analysis of Chromosome 17q markers and Breast-Ovarian Cancer in Icelandic Families, and Possible Relationship to Prostatic Cancer," *Am. J. Hum. Genet.* 52:711-717.

Bishop, T.D. and Gardner, E.J. (1980). "Analysis of the Genetic Predisposition to Cancer in Individual Pedigrees," Banbury Report 4: Cancer Incidence in Defined Populations, Cairns et al., eds., Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, pp. 389-408.

Bishop, D.T., et al. (1988). "Segregation and Linkage Analysis of Nine Utah Breast Cancer Pedigrees," *Genet. Epidemiol.* 5:151-169.

Black, D.M., et al. (1993). "A Somatic Cell Hybrid Map of the Long Arm of Human Chromosome 17, Containing the Familial Breast Cancer Locus (BRCA1)," *Am. J. Hum. Genet.* 52:702-710.

Bowcock, A.M., et al. (1993). "THRA1 and D17S183 Flank an Interval of <4cM for the Breast-Ovarian Cancer Gene (BRCA1) on Chromosome 17q21," *Am. J. Hum. Genet.* 52: 718-722.

Boyd, J. (1995). "BRCA1: More than a hereditary breast cancer gene?" *Nature Genetics* 9:335-336.

Chamberlain, J.S., et al. (1993). "BRCA1 Maps Proximal to D17S579 on Chromosome 17q21 by Genetic Analysis," *Am. J. Hum. Genet.* 52:792-798.

UNITED STATES PATENT AND TRADEMARK OFFICE
CERTIFICATE OF CORRECTION

PATENT NO. : 5,747,282
DATED : May 5, 1998
INVENTOR(S) : Mark. H. Skolnick et al.

Page 3 of 6

It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

Claus, E.B., et al. (1991). "Genetic Analysis of Breast Cancer in the Cancer and Steroid Hormone Study," *Am. J. Hum. Genet.* 48:232-242.

Cohen, B.B., et al. (1993). "Linkage of a Major Breast Cancer Gene to Chromosome 17q12-21: Results from 15 Edinburgh Families," *Am. J. Hum. Genet.* 52:723-729.

Devilee, P., et al. (1993). "Linkage to Markers for the Chromosome Region 17q12-q21 in 13 Dutch Breast Cancer Kindreds," *Am. J. Hum. Genet.* 52:730-735.

Dunphy, W.G. and Newport, J.W. (1989). "Fission Yeast p13 Blocks Mitotic Activation and Tyrosine Dephosphorylation of the *Xenopus cdc2* Protein Kinase," *Cell* 58:181-191.

Easton, D.F., et al. (1993). "Genetic Linkage Analysis in Familial Breast and Ovarian Cancer: Results form 214 Families," *Am. J. Hum. Genet.* 52:678-701.

Fain, P.R. (1992). "Third international workshop on human chromosome 17 mapping," *Cytogen. Cell Genet.* 60:178-186.

Feunteun, J., et al. (1993). "A Breast-Ovarian Cancer Susceptibility Gene Maps to Chromosome 17q21," *Am. J. Hum. Genet.* 52:736-742.

Ford, D., et al. (1993). "The risks of cancer in BRCA1 mutation carriers," *Am. J. Hum. Genet.* 53(supplement):abstract no. 298.

Futreal, P.A., et al. (1994). "BRCA1 Mutations in Primary Breast and Ovarian Carcinomas," *Science* 266:120-122.

Goldgar, D.E., et al. (1993). "Chromosome 17q Linkage Studies of 18 Utah Breast Cancer Kindreds," *Am. J. Hum. Genet.* 52:743-748

Gould, K.L. and Nurse, P. (1989). "Tyrosine Phosphorylation of the fission yeast *cdc2+* protein kinase regulates entry into mitosis," *Nature* 342:39-45.

Hall, J.M., et al. (1990). "Linkage of Early-Onset Familial Breast Cancer to Chromosome 17q21," *Science* 250:1684-1689.

Hall, J.M., et al. (1992). "Closing in on a Breast Cancer Gene on Chromosome 17q," *Am. J. Hum. Genet.* 50:1235-1241.

UNITED STATES PATENT AND TRADEMARK OFFICE
CERTIFICATE OF CORRECTION

PATENT NO. : 5,747,282
DATED : May 5, 1998
INVENTOR(S) : Mark. H. Skolnick et al.

Page 4 of 6

It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

Hosking, L., et al. (1995). "A somatic *BRCA1* mutation in an ovarian tumour," *Nature Genetics* 9:343-344.

Ishibashi, T., et al. (1992). "Expression cloning of a human dual-specificity phosphatase," *Proc. Natl. Acad. Sci. USA* 89:12170-12174.

Kamb, A., et al. (1994). "Localization of the VHR Phosphatase Gene and Its Analysis as a Candidate for *BRCA1*," *Genomics* 23: 163-167.

Kumagai, A. and Dunphy, W.G. (1991). "The *cdc25* Protein Controls Tyrosine Dephosphorylation of the *cdc2* Protein in a Cell-Free system," *Cell* 64:903-914.

Lindblom, A., et al. (1993). "Linkage Analysis with Markers on 17q in 29 Swedish Breast Cancer Families," *Am. J. Hum. Genet.* 52:749-753.

Malkin, D., et al. (1990). "Germ Line *p53* Mutations in a familial Syndrome of Breast Cancer, Sarcomas, and Other Neoplasms," *Science* 250:1233-1238.

Margaritte, P., et al. (1992). "Linkage of Familial Breast Cancer to Chromosome 17q21 May Not be Restricted to Early-Onset Disease," *Am. J. Hum. Genet.* 50: 1231-1234.

Mazoyer, S., et al. (1993). "Linkage Analysis of 19 French Breast Cancer Families, with Five Chromosome 17q Markers," *Am. J. Hum. Genet.* 52:754-760.

Merajver, S.D., et al. (1995). "Somatic mutation s in the *BRCA1* gene in sporadic ovarian tumours," *Nature Genetics* 9:439-443.

Miki, Y., et al. (1994). "A Strong Candidate for the Breast and Ovarian Cancer Susceptibility Gene *BRCA1*," *Science* 266:66-71.

Milner, B.J., et al. (1993). "Linkage "Studies with 17q and 18q Markers in a Breast/Ovarian Cancer Family," *Am. J. Hum. Genet.* 52:761-766.

Narod, S.A., et al. (1991). "Familial breast-ovarian cancer locus on chromosome 17q12-q23," *Lancet* 338:82-83.

Newman, B. et al. (1988). "Inheritance of human breast cancer: Evidence for autosomal dominant transmission in high-risk families," *Proc. Natl. Acad. Sci. USA* 85:3044-3048.

UNITED STATES PATENT AND TRADEMARK OFFICE
CERTIFICATE OF CORRECTION

PATENT NO. : 5,747,282
DATED : May 5, 1998
INVENTOR(S) : Mark. H. Skolnick et al.

Page 5 of 6

It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

O'Connell, P., et al. (1994). "A Radiation Hybrid Map of the BRCA1 Region," *Am. J. Hum. Genet.* 54:526-534.

Roberts, L. (1993). "Zeroing In on a Breast Cancer Susceptibility Gene," *Science* 259:622-625.

Shattuck-Eidens, D., et al. (1995). "A Collaborative Survey of 80 Mutations in the BRCA1 Breast and Ovarian Cancer Susceptibility Gene." *JAMA* 273:535-541.

Simard, J., et al. (1994). "Common origins of BRCA1 mutations in Canadian breast and ovarian cancer families." *Nature Genetics* 8:392-398.

Smith, S.A., et al. (1992). "Allele losses in the region 17q12-21 in familial breast and ovarian cancer involve the wild-type chromosome," *Nature Genetics* 2:128-131.

Smith, S.A.; et al. (1993). "Genetic Heterogeneity and Localization of a Familial Breast-Ovarian Cancer Gene on Chromosome 17q12-q21," *Am. J. Hum. Genet.* 52:767-776.

Spurr, N.K., et al. (1993). "Linkage Analysis of Early-Onset Breast and Ovarian Cancer Families, with Markers on the Long Arm of Chromosome 17," *Am. J. Hum. Genet.* 52:777-785.

Teare, M.D., et al. (1993). "A Linkage Study in Seven Breast Cancer Families," *Am. J. Hum. Genet.* 52:786-788.

Thompson, M.E., et al. (1995). "Decreased expression of *BRCAl* accelerates growth and is often present during sporadic breast cancer progression," *Nature Genetics* 9: 444-450.

Ullrich, A. and Schlessinger, J. (1990). "Signal Transduction by Receptors with Tyrosine Kinase Activity," *Cell* 61:203-212.

Williams, W.R. and Anderson, D.E. (1984). "Genetic Epidemiology of Breast Cancer: Segregation Analysis of 200 Danish Pedigrees." *Genet. Epidemiol.* 1:7-20.

Zimmerman, W., et al. (1993). "Linkage Analysis in German Breast Cancer Families with Early Onset of the Disease, Using Highly Polymorphic Markers from the Chromosome 17q11-q24 Region." *Am. J. Hum. Genet.* 52:789-791

UNITED STATES PATENT AND TRADEMARK OFFICE
CERTIFICATE OF CORRECTION

PATENT NO. : 5,747,282
DATED : May 5, 1998
INVENTOR(S) : Mark. H. Skolnick et al.

Page 6 of 6

It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

Go, R.C.P., et al. (1983). "Genetic Epidemiology of Breast Cancer and Associated Cancers in High-Risk Families. I. Segregation Analysis," *J. Natl. Cancer Inst.* 71:455-461. --;

Column 36,

Line 36, "collaborators" should be -- collaborators' --;

Column 155,

Line 17, "saod" should be -- said --.

Signed and Sealed this

Fourth Day of December, 2001

Attest:

Nicholas P. Godici

Attesting Officer

NICHOLAS P. GODICI
Acting Director of the United States Patent and Trademark Office

A000248

UNITED STATES PATENT AND TRADEMARK OFFICE
CERTIFICATE OF CORRECTION

PATENT NO. : 5,747,282
APPLICATION NO. : 08/483554
DATED : May 5, 1998
INVENTOR(S) : Skolnick et al.

Page 1 of 1

It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

Col. 155, Claim 16, line 24, please change "nuycleotide" to --nucleotide--.

Col. 155, Claim 16, line 25, please change "chin" to --chain--.

Col. 156, Claim 18, line 4, please change "ovariann" to --ovarian--.

Col. 156, Claim 18, line 5, please change "prime" to --primer--.

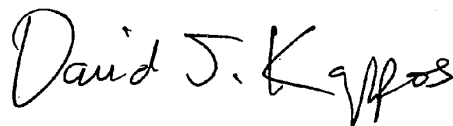
Col. 156, Claim 18, line 6, please change "detectiong" to --detecting--.

Col. 156, Claim 19, line 10, please change "sisceptibility" to --susceptibility--.

Col. 156, Claim 19, line 11, please change "oligimucleotide" to --oligonucleotide--.

Signed and Sealed this

Twenty-fourth Day of November, 2009



David J. Kappos
Director of the United States Patent and Trademark Office



US005753441A

United States Patent [19]
Skolnick et al.

[11] **Patent Number:** **5,753,441**
[45] **Date of Patent:** **May 19, 1998**

- [54] **170-LINKED BREAST AND OVARIAN CANCER SUSCEPTIBILITY GENE**
- [75] **Inventors:** **Mark H. Skolnick; David E. Goldgar; Yoshio Miki; Jeff Swenson; Alexander Kamb; Keith D. Harshman; Donna M. Shattuck-Eidens; Sean V. Tavtigian**, all of Salt Lake City, Utah; **Roger W. Wiseman; P. Andrew Futreal**, both of Durham, N.C.
- [73] **Assignees:** **Myriad Genetics, Inc.; University of Utah Research Foundation**, both of Salt Lake City, Utah; **The United States of America as represented by the Department of Health and Human Services**, Washington, D.C.

- [21] **Appl. No.:** **488,011**
[22] **Filed:** **Jan. 5, 1996**

Related U.S. Application Data

- [63] Continuation-in-part of Ser. No. 409,305, Mar. 24, 1995, abandoned, which is a continuation-in-part of Ser. No. 348,824, Nov. 29, 1994, abandoned, which is a continuation-in-part of Ser. No. 308,104, Sep. 16, 1994, abandoned, which is a continuation-in-part of Ser. No. 300,266, Sep. 2, 1994, abandoned, which is a continuation-in-part of Ser. No. 289,221, Aug. 12, 1994, abandoned.
- [51] **Int. Cl.⁶** **C12Q 1/68; C12P 19/34; C07H 21/02; C07H 21/04**
- [52] **U.S. Cl.** **435/6; 435/91.1; 435/91.2; 435/7.1; 435/7.2; 435/7.9; 435/4; 436/548; 436/500; 530/387.2; 530/388.1; 424/1.11; 424/88; 536/23.1; 536/24.3; 536/24.33**
- [58] **Field of Search** **435/6, 91.1, 91.2, 435/7.1-7.9, 4; 536/23.1, 24.3-24.33, 548; 436/500; 424/1.11, 88; 530/387.2, 388.1**

[56] **References Cited**

U.S. PATENT DOCUMENTS

5,236,844 8/1993 Basset et al. .

FOREIGN PATENT DOCUMENTS

0 518 650 12/1992 European Pat. Off. .
91/09964 7/1991 WIPO .
92/00311 1/1992 WIPO .
92/11874 7/1992 WIPO .
94/00764 1/1994 WIPO .
95/19369 7/1995 WIPO .

OTHER PUBLICATIONS

Friedman, L.S. et al. (1994). "Confirmation of BRCA1 by analysis of germline mutations linked to breast and ovarian cancer in ten families." *Nature Genetics* 8:399-404.

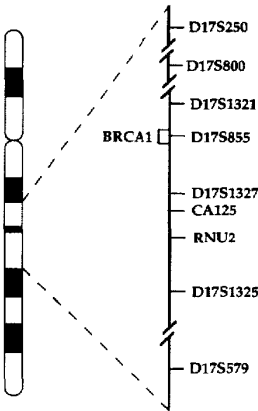
(List continued on next page.)

Primary Examiner—W. Gary Jones
Assistant Examiner—Dianne Rees
Attorney, Agent, or Firm—Venable, Baetjer, Howard & Civiletti, LLP

[57] **ABSTRACT**

The present invention relates generally to the field of human genetics. Specifically, the present invention relates to methods and materials used to isolate and detect a human breast and ovarian cancer predisposing gene (BRCA1), some mutant alleles of which cause susceptibility to cancer, in particular breast and ovarian cancer. More specifically, the invention relates to germline mutations in the BRCA1 gene and their use in the diagnosis of predisposition to breast and ovarian cancer. The present invention further relates to somatic mutations in the BRCA1 gene in human breast and ovarian cancer and their use in the diagnosis and prognosis of human breast and ovarian cancer. Additionally, the invention relates to somatic mutations in the BRCA1 gene in other human cancers and their use in the diagnosis and prognosis of human cancers. The invention also relates to the therapy of human cancers which have a mutation in the BRCA1 gene, including gene therapy, protein replacement therapy and protein mimetics. The invention further relates to the screening of drugs for cancer therapy. Finally, the invention relates to the screening of the BRCA1 gene for mutations, which are useful for diagnosing the predisposition to breast and ovarian cancer.

37 Claims, 18 Drawing Sheets

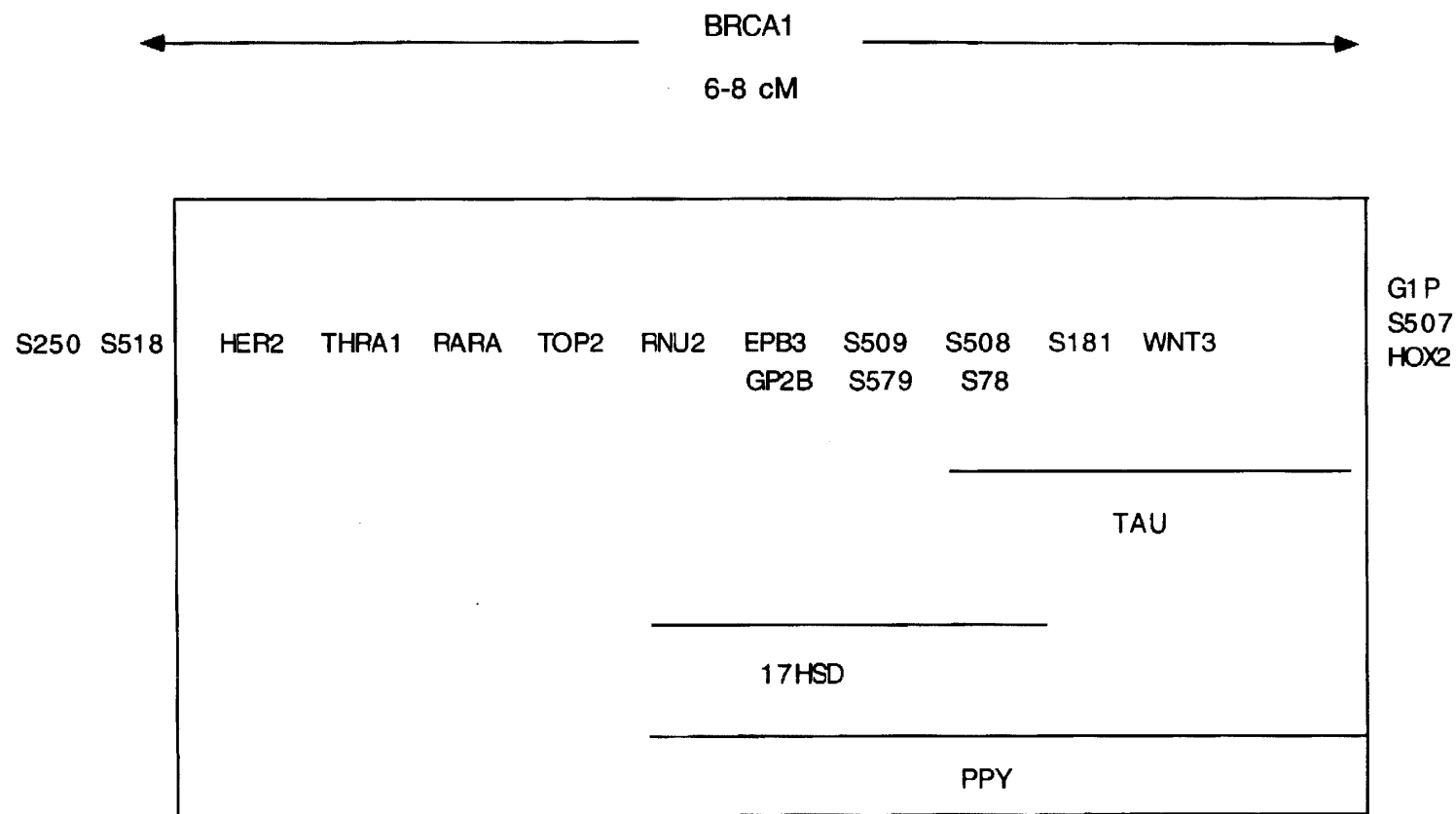


5,753,441

Page 3

-
- Spurr, N.K., et al. (1993). "Linkage Analysis of Early-Onset Breast and Ovarian Cancer Families, with Markers on the Long Arm of Chromosome 17." *Am. J. Hum. Genet.* 52:777-785.
- Teare, M.D., et al. (1993). "A Linkage Study in Seven Breast Cancer Families." *Am. J. Hum. Genet.* 52: 786-788.
- Thompson, M.E., et al. (1995). "Decreased expression of BRCA1 accelerates growth and is often present during sporadic breast cancer progression." *Nature Genetics* 9:444-450.
- Ullrich, A. and Schlessinger, J. (1990). "Signal Transduction by Receptors with Tyrosine Kinase Activity." *Cell* 61:203-212.
- Williams, W.R. and Anderson, D.E. (1984). "Genetic Epidemiology of Breast Cancer: Segregation Analysis of 200 Danish Pedigrees." *Genet. Epidemiol.* 1:7-20.
- Zimmermann, W., et al. (1993). "Linkage Analysis in German Breast Cancer Families with Early Onset of the Disease, Using Highly Polymorphic Markers from the Chromosome 17q11-q24 Region." *Am. J. Hum. Genet.* 52:789-791.
- Goldberg et al. *Clinical Chemistry* 39: 2360-2374, 1993.
- Wooster et al., *Science* 265:2088-2090, 1994.
- Sato et al., *Cancer Res.* 52: 1643-1646, 1992.
- Malkin et al., *Science* 250: 1233-1238, 1990.
- Boyd, *Nature Genetics* 9: 335-336, 1995.
- Merajver et al., *Nature Genetics* 9:439-443, 1995.
- Castilla et al *Nature Genetics* 8: 387-391, 1994.
- Friedman et al., *Nature Genetics* 8: 1-6, 1994.
- Goldgar et al. *J. of Nat. Cancer Inst.* 86: 200-209, 1994.

A000252



Map of the early onset breast and ovarian cancer region (BRCA1)

FIG. 1

A000253

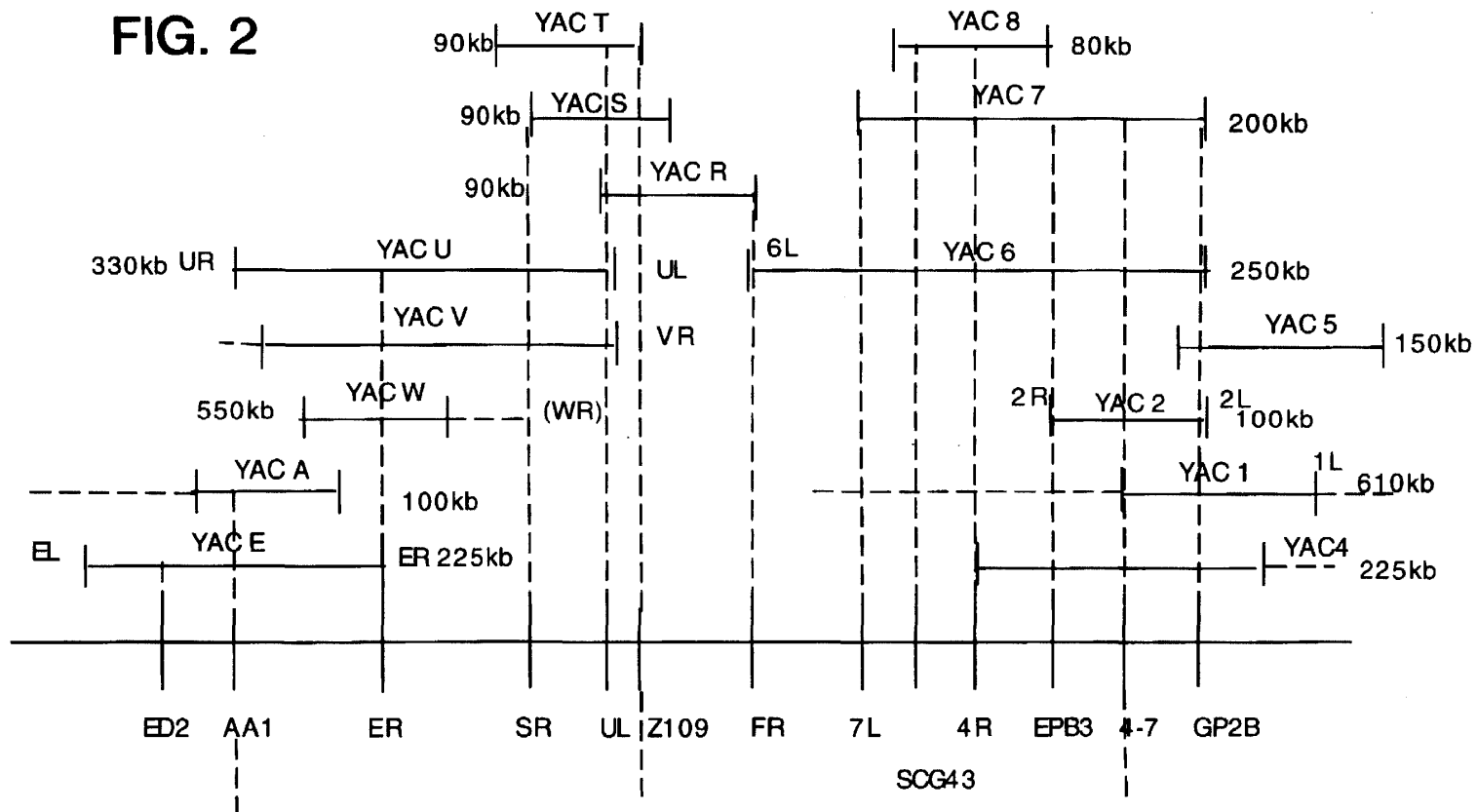
U.S. Patent

May 19, 1998

Sheet 1 of 18

5,753,441

FIG. 2



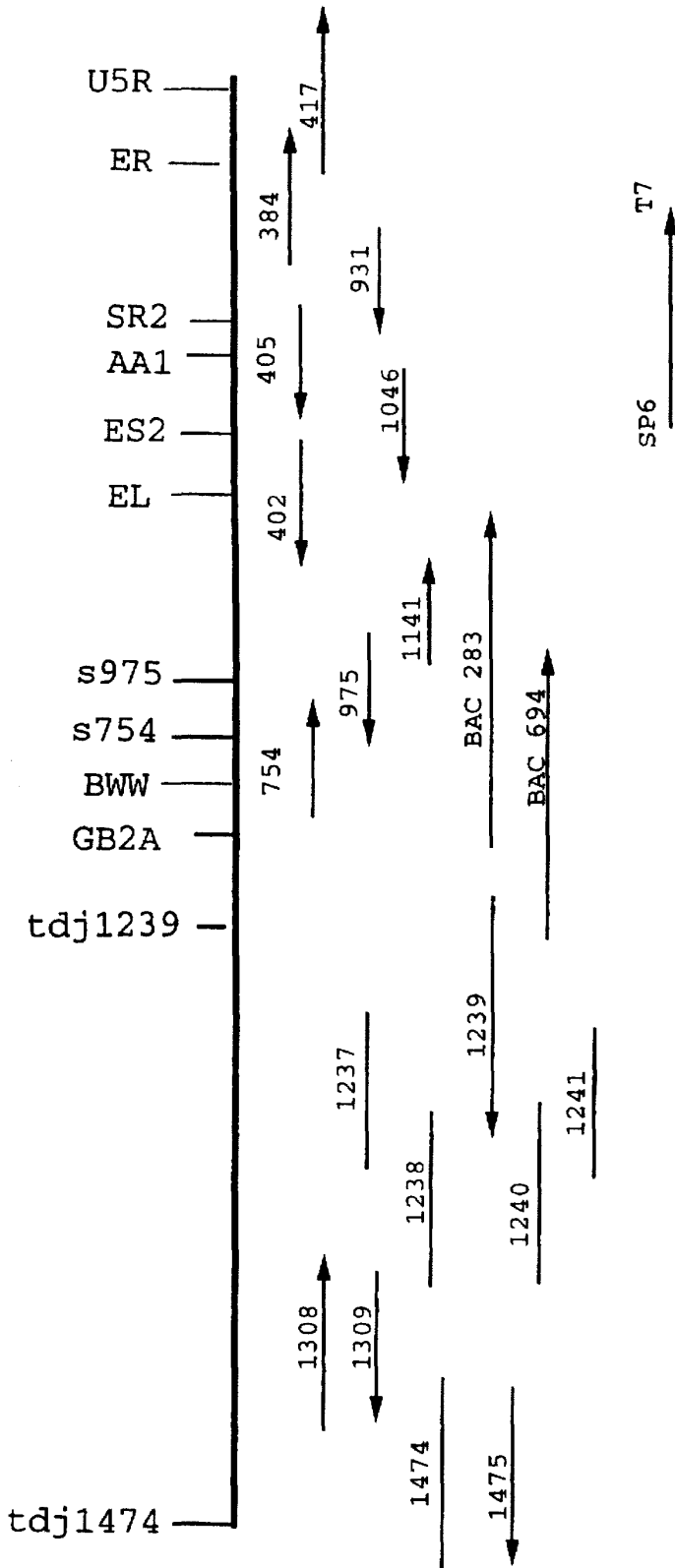
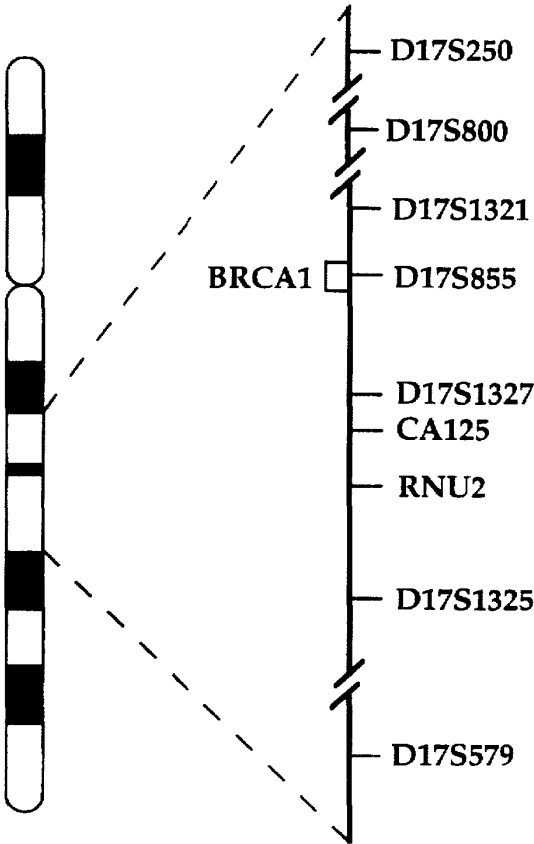


FIG. 3

FIG. 4



SEQ. ID
NO:

82 BRCA1
83 RPT1
84 RIN1
85 RFP1
C3HC4 motif

CPICLELIKEPVSTK-CDHIFCKFCMLKLLNQKK---GPSQCPLCK
CPICLELLKEPVSAD-CNHSFCRACITLNYESNRNTDGKGNCPVCR
CPICLDMLKNTMTTKECLHRFCSDCIVTALRS-----GNKECPTCR
CPVCLQYFAEPMMLD-CGHNICCACLARCWGTAC---TNVSCPQCR
C--C-----C-H--C--C-----C--C

FIG. 5

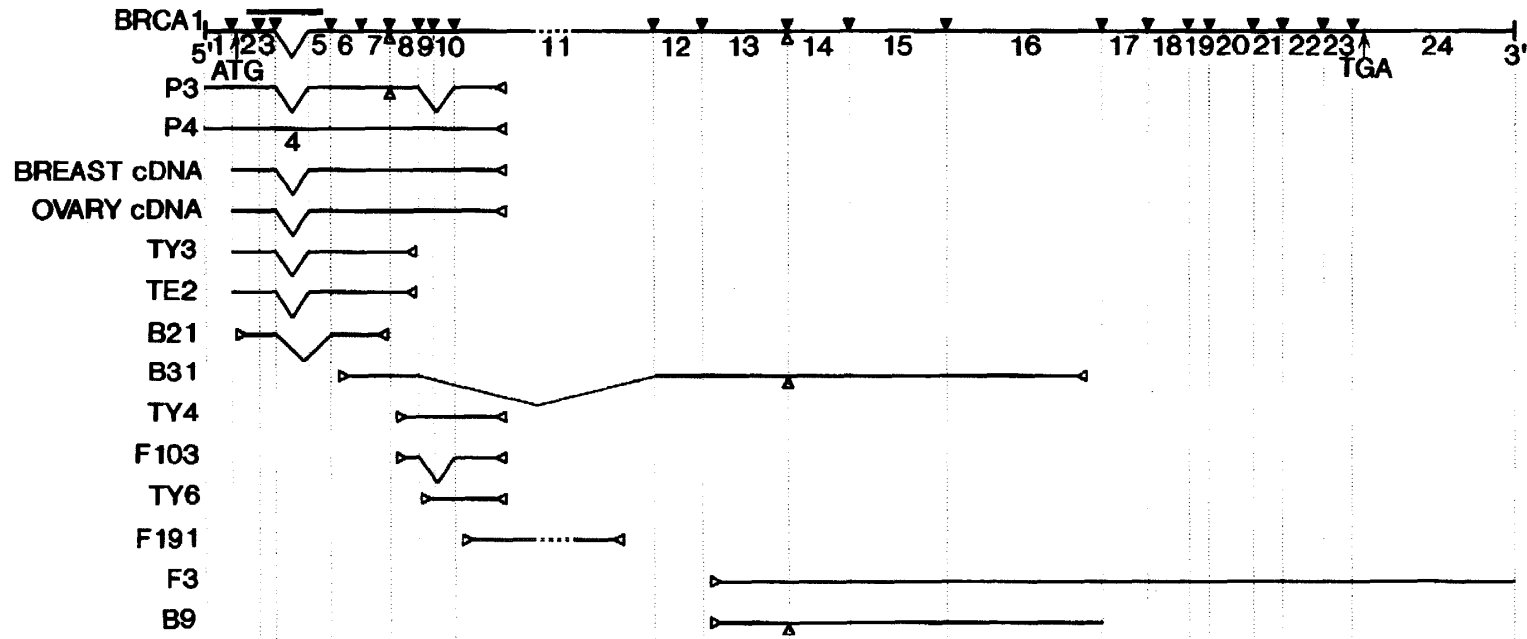


FIG. 6

U.S. Patent

May 19, 1998

Sheet 5 of 18

5,753,441

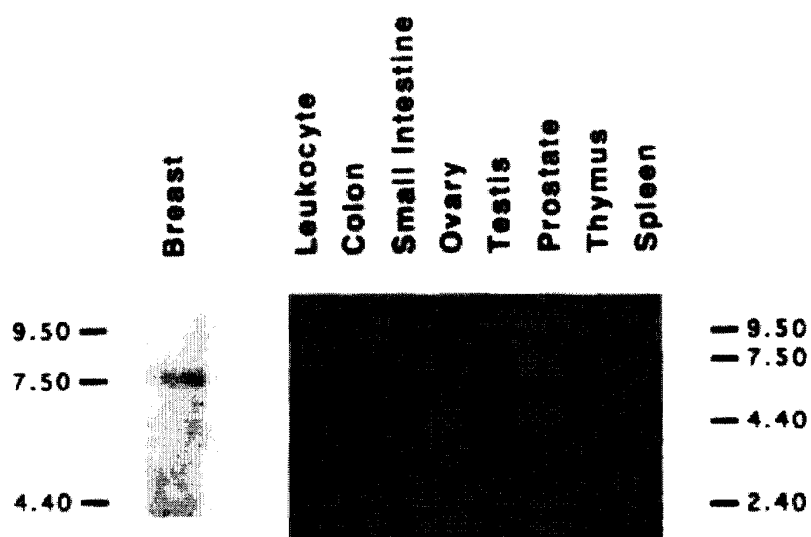


FIG. 7

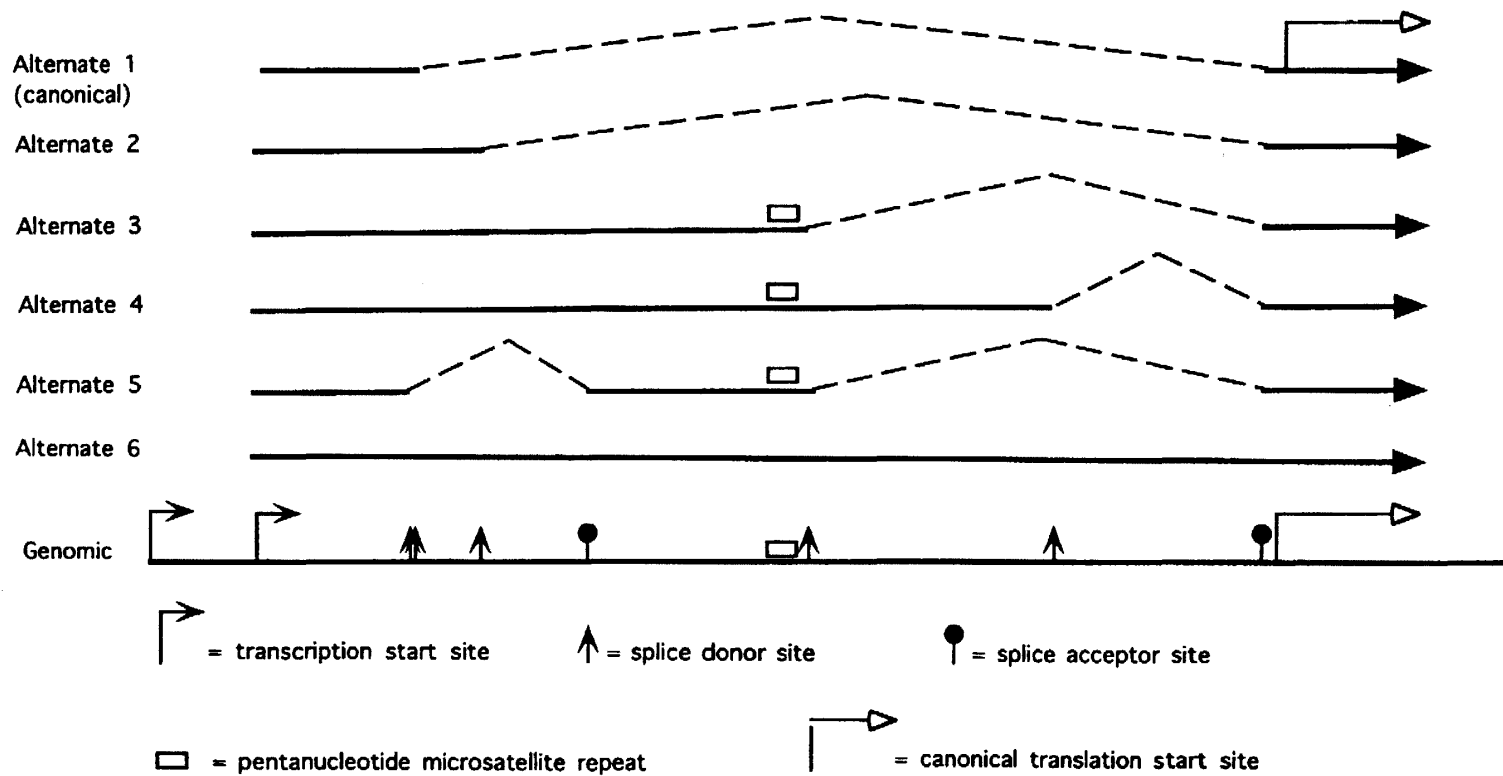


FIG. 8

U.S. Patent

May 19, 1998

Sheet 7 of 18

5,753,441

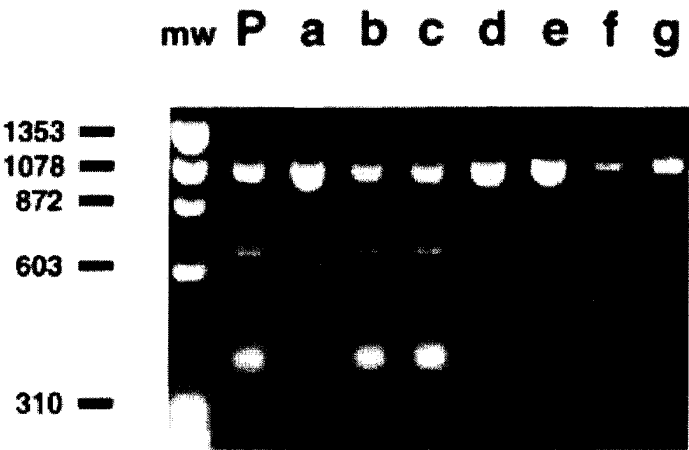


FIG. 9A

U.S. Patent

May 19, 1998

Sheet 9 of 18

5,753,441

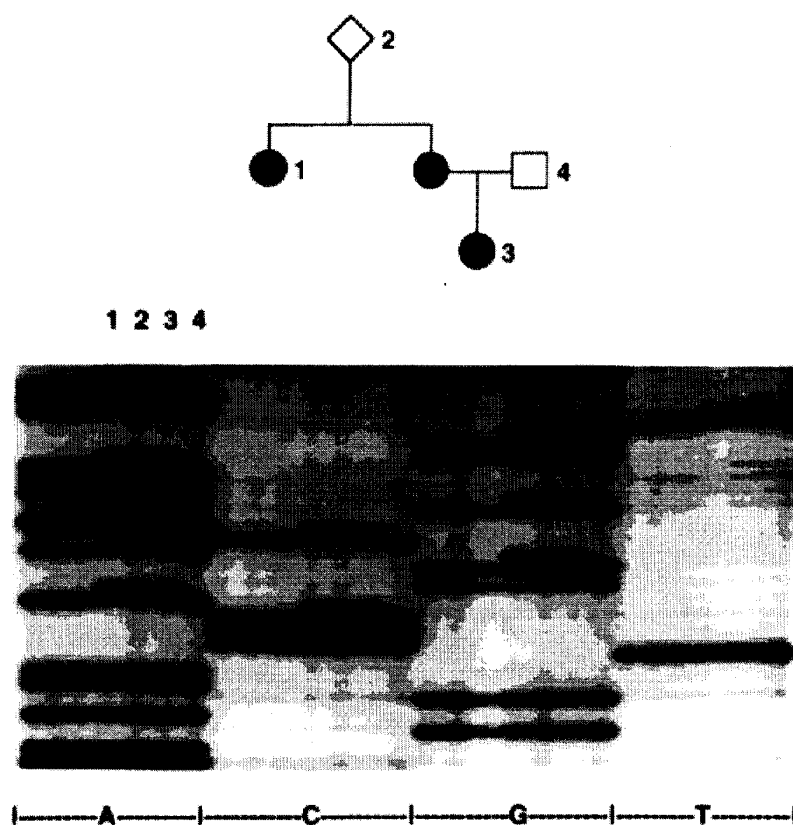
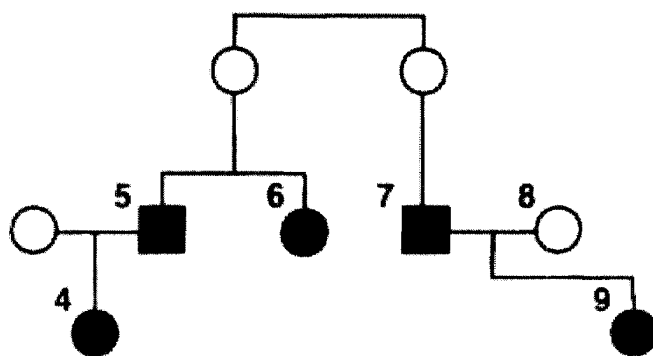
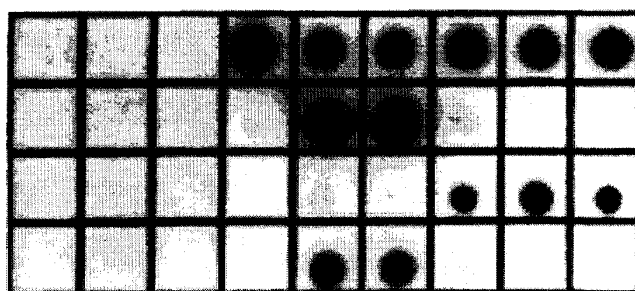


FIG. 9B



PM1

4 5 6 7 8 9



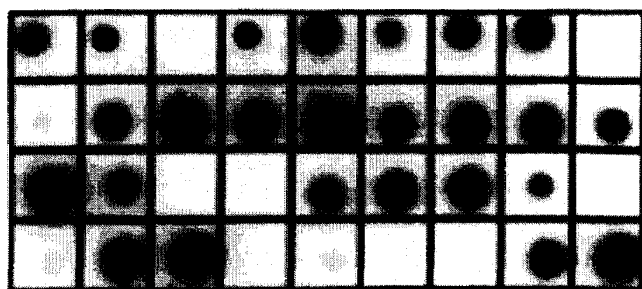
A

G

A

G

PM7



87

G

A

G

FIG. 9C

A000263

A000264

A000265

A000266

A000267

15301 gaattttaactcattttgtgtgacatgaaagtaaatccagtcctgccaatgagaagaaaaag
15361 acacagcaagttgcagcgtttatagtctgcttttacatctgaacctctgttttggttatt
15421 taagGTGAAGCAGCATCTGGGTGTGAGAGTGAAACAAGCGTCTCTGAAGACTGCTCAGGG
15481 CTATCCTCTCAGAGTGACATTTTAACCACCTCaggtaaaaagcgtgtgtgtgtgtgcacat
15541 gcgtgtgtgtgggtgtcctttgcattcagtagtatgtatcccacattcttaggtttgtcta
15601 catcatctcctttgaattaatggcacaaattgtttgtgggttcatttgtcvvvvvvvvvvvn
15661 gngaattgtaatcctaataatttncnccnacttaaagaataaccactccaanggcacnca
15721 atacatcaatcaattggggaattgggattttccctcnctaacaatcantggaataatttca
15781 tggcattaattgcatgaaatgtgggttagattanaagggtgttcatgctagaacttgtagtte
15841 catactagggtgatttcaattcctgtgctaaaattaaatttgatatgatattntcatttaa
15901 tggaaagcttctcaaagtatttcattttcttgggtaccatttatcgtttttgaAGCAGAGG
15961 GATACCATCAACATAACTGATAAAAGCTCCAGCAGGAAATGGCTGAACTAGAAGCTGTG
16021 TTAGAACAGCATGGGAGCCAGCCTTCTAACAGCTACCCTTCATCATAGTGACTCTTCT
16081 GCCCTTGAGGACCTGCAGAAATCCAGAACAAAGCACATCAGAAAAAGgtgtgtattgttgg
16141 ccaaacactgatatacttaagcaaaattcttctccttccccctttatctccttctgaagagta
16201 aggacctagctccaacattttatgatccttgctcagcacatgggtaattatggagccttg
16261 gttcttgtccctgctcacaactaataataccagtcagaggggaccaaggcagtcattcatg
16321 ttgtcatctgagatacctacaacaagtagatgctatggggagcccatggvvvvvvvvvvv
16381 vvcattgggtgctagcatctgtctgttgcatgtcttgtgtttataaaaattctgcctgata
16441 tacttgttaaaaaccaatttgtgtatcatagattgatgcttttgaaaaaaaaatcagtattc
16501 taacctgaattatcactatcagaacaaagcagtaaaagtagatttgttttctcattccatt
16561 taaagCAGTATTAAC TTCACAGAAAAGTAGTGAATACCCTATAAGCCAGAATCCAGAAGG
16621 CCTTCTGCTGACAAGTTTGAGGTGTCTGCAGATAGTTCTACCAGTAAAAATAAAGAACC
16681 AGGAGTGGAAGGtaagaacatcaatgtaaaagatgctgtgggtatctgacatctttattt
16741 atattgaaactctgattgttaatttttttaccatactttctccagtttttttgcatacvg
16801 gcatttatacatttttattgtctctaggatacttcttttgttttaactctatataggvvvvv
16861 vvvvvvvvggataagnccaagagatattttgatagggtgagcagtgnaattngnaaaa
16921 ttnctgcctgcttttaattcttcccccgcttcttcttccctnccctccctccctnccct
16981 cccgctcctncccttcttcttccctccctccctccctncccttcttctnctcttcttctt
17041 tccgtgtctaccttcttcttccctccctccctcccttcttcttcttcttcttcttctt
17101 ctttcccttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt
17161 agtgcagtgggcgtgatctcgncctactgcaacctctgtctcccaggttcaagcaatttct
17221 ctgcctcagcctcccgagtagctgagattacaggcgccagccaccacccagctactga
17281 cctgcttttvvvvvvvvvvvvaacagctgggagatattgggtgcctcagaccaaaccccat
17341 gtatatgtcaaccttgacatatgggcaggcaacatgaatccagacttctagggtgtcat
17401 gccgggtcttttttggcagtcatttctgatctctctgacatgagctgtttcatttatgct
17461 ttggctgccccaagtagatttgtccttttcaaatgggtggcgatggtttctctctt
17521 catttatctttctagTCACTCCCCCTCTAAATGCCCATCATTAGATGATAGGTGGTACAT
17581 GCACAGTTGCTCTGGGAGCTCTTCAGAAATGAGAAATACCCATTCTCAAGAGGAGCTCATTAA
17641 GGTTGTTGATGTGGAGGAGCAACAGCTGGAAGAGTCTGGGCCACACGATTTGACGGAAAC
17701 ATCTTACTTGCCAAGGCAAGATCTAGgtaataatttcatctgctgtattggaacaaacact
17761 ytgattttactctgaaatcctacataaagataattctgggttaaccaacttttagatgtacta
17821 gtctatcatggacacttttgttatacttaattaaagccactttagaaaaatagctcaagt
17881 gttaatcaagggttacttgaaaaattattgaaactgttaatccatctatattttaattaat
17941 ggtttaactaatgattttgaggatgwgggagctcktggtgtactctamatgtattatttca
18001 ggccaggcatagtggtcacgcctggtaatcccagtayycmrgagcccgaggcgaggtgga
18061 gccagctgaggtcaggagttcaagacctgtcttggccaacatggngaaacctgtcttct
18121 ttcttaaaaaanacaaaaaaattaaactgggttgtgcttaggtgnatgccccgnatccta
18181 gttnttcttgnnggttgaggaggagatcacnttggaccceggaggggnggggtggggng
18241 agcaggncaaaaacnacagccagctgggtgggaagggaagcccatcnaaaaaanntnv
18301 vvvvvvvvvvtttttaggaacaagctacttttqattttccaccaaacctgtattcat

FIG. 10F

A000269

5,753,441

21421 acctcaagtgatctgccagcctcagttctcccaaagtgtaggattacagggggtgagcac
21481 tgcgcctggcctgaatgcctaaaatatgacgtgtctgctccacttccattgaaggaagct
21541 tctctttctcttatctgatgggttgtgtttgggtttcttttcagCATGATTTTGAAGTCAG
21601 AGGAGATGTGGTCAATGGAAGAAACCACCAAGGTCCAAAGCGAGCAAGAGAATCCCAGGA
21661 CAGAAAAGgtaaagctccctccctcaagttgacaaaaatctcaccaccactctgtattc
21721 cactcccccttgcagagatgggcccgttcatlttgtaagacttattacatacatacacag
21781 tgctagatactttcacacagggttcttttttctactcttccatcccaaccacataaataagt
21841 attgtctctactttatgaatgataaaaactaagagatttagagaggctgtgtaatttgga
21901 tcccgctctcggttcagatcvvvvvvvvvvvtggcctgattggtgacaaaagtgaga
21961 tgctcagtccttgaatgacaaagaatgcctgtagagttgcagggtccaactacatatgcac
22021 ttcaagaagatcttctgaaatctagtagtgttctggacattggactgcttgtccctggga
22081 agtagcagcagaaatgatcggtggtgaacagaagaaaaagaaagctcttctcttttgaa
22141 agtctgttttttgataaaaagccaatattcttttataactagatttctctctctccatt
22201 cccctgtccctctctcttctctctcttcttccagATCTTCAGGGGGCTAGAAATCTGTTGC
22261 TATGGGCCCTTCACCAACATGCCCCACAGttaagagcctgggagaacccagagttccagc
22321 accagcctttgtcttacatagtggagtattataagcaagggtcccacgatgggggttctc
22381 agattgctgaaatgttctagaggctattctatttctctaccactctccaaacaaacagc
22441 acctaattgttatcttatggcaaaaaaaaaactataaccttctccccctctcaagagctg
22501 aagggtggttaattaggttaggattcagtatgttatgtgttcagatggcgttcagctgctgt
22561 agtgcvvvvvvvvvvvvvttgagagactatcaaaccttataccaagtggccttatgga
22621 gactgataaccagagtacatggcatatcagtggcaaattgacttaaaatccataccctta
22681 ctatttttaagaccattgtcctttgggagcagagagacagactctccattgagaggctctg
22741 ctataagccttcatcgggagagtgtagggtagagggcctgggttaagtatgcagattact
22801 gcagtgattttacatgtaaatgtccatttttagATCAACTGGAATGGATGGTACAGCTGTG
22861 TGGTGCTTCTGTGGTGAAGGAGCTTTTCATTACACCTTGGCACAgtaagtattgggtg
22921 ccctgtcagtggtggaggacacaatatctctcctgtgagcaagactggcacctgtcagt
22981 ccctatggatgccctactgtagcctcagaagtcttctctgcccacataacctgtgccaaa
23041 agactccatvvvvvvvvvvvggttggtacgtgtctgtagtccagctacttgggaggct
23101 gagatggaaggattgcttgagcccaggaggcagagggtggnannttacgctgagatcacac
23161 cactgcactccagcctgggtgacagagcaagaccctgtctcaaaaacaaacaaaaaaat
23221 gatgaagtgcaggtccagtagtctactttgacactttgaatgctcttccctcctggg
23281 gatccagGGTGTCCACCCAATTGTGGTTGTGCAGCCAGATGCCTGGACAGAGGACAATGG
23341 CTTCATGgtaaggtgcctcgcatgtacctgtgctatttagtgggtcctgtgcatgggt
23401 ttgggttatcactcattacctgggtgcttgagtagcacagttcttggcacatttttaaata
23461 tttgttgaatgaatggctaaaatgtctttttgatgtttttattgtttattgttttatatt
23521 gtaaaagtaatacatgaactgtttccatgggggtgggagtaagatatgaatgttcatcacv
23581 vvvvvvvvvvvvcagtaatccnagaactcatacgaccgggcccctggagtcgntgnttn
23641 gagectagtcnngagaatgaattgacactaatctctgcttgtgttctctgtctccagCA
23701 ATTGGGCAGATGTGTGAGGCACCTGTGGTGACCCGAGAGTGGGTGTGGACAGTGTAGCA
23761 CTCTACCAGTGCCAGGAGCTGGACAGCCTACCTGATACCCCAGATCCCCACAGCCACTAC
23821 TGACTGCAGCCAGCCACAGGTACAGAGCCACAGGACCCCCAAGATGAGCTTACAAAGTGG
23881 CCTTTCCAGGCCCTGGGAGCTCCTCTCAGTCTCTACTGTCTCTGGCTACTAA
23941 ATATTTTTATGTACATCAGCCTGAAAAGGACTTCTGGCTATGCAAGGGTCCCTTAAAGATT
24001 TTCTGCTTGAAGT(TCCCTTGGAAAT

FIG. 10H

5,753,441

1

170-LINKED BREAST AND OVARIAN CANCER SUSCEPTIBILITY GENE

CROSS REFERENCE TO RELATED APPLICATIONS

This application is a continuation-in-part of application Ser. No. 08/409,305 filed on 24 March 1995 now abandoned, which is a continuation-in-part of application Ser. No. 08/348,824 filed on 29 November 1994 now abandoned, which is a continuation-in-part of application Ser. No. 08/308,104 filed on 16 September 1994 now abandoned, which is a continuation-in-part of application Ser. No. 08/300,266, filed on 2 September 1994 now abandoned, which is a continuation-in-part of application Ser. No. 08/289,221, filed on 12 August 1994 now abandoned, all incorporated herein by reference.

FIELD OF THE INVENTION

The present invention relates generally to the field of human genetics. Specifically, the present invention relates to methods and materials used to isolate and detect a human breast and ovarian cancer predisposing gene (BRCA1), some mutant alleles of which cause susceptibility to cancer, in particular, breast and ovarian cancer. More specifically, the invention relates to germline mutations in the BRCA1 gene and their use in the diagnosis of predisposition to breast and ovarian cancer. The present invention further relates to somatic mutations in the BRCA1 gene in human breast and ovarian cancer and their use in the diagnosis and prognosis of human breast and ovarian cancer. Additionally, the invention relates to somatic mutations in the BRCA1 gene in other human cancers and their use in the diagnosis and prognosis of human cancers. The invention also relates to the therapy of human cancers which have a mutation in the BRCA1 gene, including gene therapy, protein replacement therapy and protein mimetics. The invention further relates to the screening of drugs for cancer therapy. Finally, the invention relates to the screening of the BRCA1 gene for mutations, which are useful for diagnosing the predisposition to breast and ovarian cancer.

The publications and other materials used herein to illuminate the background of the invention, and in particular, cases to provide additional details respecting the practice, are incorporated herein by reference, and for convenience, are referenced by author and date in the following text and respectively grouped in the appended List of References.

BACKGROUND OF THE INVENTION

The genetics of cancer is complicated, involving multiple dominant, positive regulators of the transformed state (oncogenes) as well as multiple recessive, negative regulators (tumor suppressor genes). Over one hundred oncogenes have been characterized. Fewer than a dozen tumor suppressor genes have been identified, but the number is expected to increase beyond fifty (Knudson, 1993).

The involvement of so many genes underscores the complexity of the growth control mechanisms that operate in cells to maintain the integrity of normal tissue. This complexity is manifest in another way. So far, no single gene has been shown to participate in the development of all, or even the majority of human cancers. The most common oncogenic mutations are in the *II-ras* gene, found in 10–15% of all solid tumors (Anderson et al., 1992). The most frequently mutated tumor suppressor genes are the TP53 gene, homozygously deleted in roughly 50% of all tumors, and

2

CDKN2, which was homozygously deleted in 46% of tumor cell lines examined (Kamb et al., 1994). Without a target that is common to all transformed cells, the dream of a “magic bullet” that can destroy or revert cancer cells while leaving normal tissue unharmed is improbable. The hope for a new generation of specifically targeted antitumor drugs may rest on the ability to identify tumor suppressor genes or oncogenes that play general roles in control of cell division.

The tumor suppressor genes which have been cloned and characterized influence susceptibility to: 1) Retinoblastoma (RB1); 2) Wilms’ tumor (WT1); 3) Li-Fraumeni (TP53); 4) Familial adenomatous polyposis (APC); 5) Neurofibromatosis type 1 (NF1); 6) Neurofibromatosis type 2 (NF2); 7) von Hippel-Lindau syndrome (VHL); 8) Multiple endocrine neoplasia type 2A (MEN2A); and 9) Melanoma (CDKN2).

Tumor suppressor loci that have been mapped genetically but not yet isolated include genes for:

Multiple endocrine neoplasia type 1 (MEN1); Lynch cancer family syndrome 2 (LCFS2); Neuroblastoma (NB); Basal cell nevus syndrome (BCNS); Beckwith-Wiedemann syndrome (BWS); Renal cell carcinoma (RCC); Tuberous sclerosis 1 (TSC1); and Tuberous sclerosis 2 (TSC2). The tumor suppressor genes that have been characterized to date encode products with similarities to a variety of protein types, including DNA binding proteins (WT1), ancillary transcription regulators (RB1), GTPase activating proteins or GAPs (NF1), cytoskeletal components (NF2), membrane bound receptor kinases (MEN2A), cell cycle regulators (CDKN2) and others with no obvious similarity to known proteins (APC and VHL).

In many cases, the tumor suppressor gene originally identified through genetic studies has been shown to be lost or mutated in some sporadic tumors. This result suggests that regions of chromosomal aberration may signify the position of important tumor suppressor genes involved both in genetic predisposition to cancer and in sporadic cancer.

One of the hallmarks of several tumor suppressor genes characterized to date is that they are deleted at high frequency in certain tumor types. The deletions often involve loss of a single allele, a so-called loss of heterozygosity (LOH), but may also involve homozygous deletion of both alleles. For LOH, the remaining allele is presumed to be nonfunctional, either because of a preexisting inherited mutation, or because of a secondary sporadic mutation.

Breast cancer is one of the most significant diseases that affects women. At the current rate, American women have a 1 in 8 risk of developing breast cancer by age 95 (American Cancer Society, 1992). Treatment of breast cancer at later stages is often futile and disfiguring, making early detection a high priority in medical management of the disease. Ovarian cancer, although less frequent than breast cancer is often rapidly fatal and is the fourth most common cause of cancer mortality in American women. Genetic factors contribute to an ill-defined proportion of breast cancer incidence, estimated to be about 5% of all cases but approximately 25% of cases diagnosed before age 40 (Claus et al., 1991). Breast cancer has been subdivided into two types, early-age onset and late-age onset, based on an inflection in the age-specific incidence curve around age 50. Mutation of one gene, BRCA1, is thought to account for approximately 45% of familial breast cancer, but at least 80% of families with both breast and ovarian cancer (Easton et al., 1993).

Intense efforts to isolate the BRCA1 gene have proceeded since it was first mapped in 1990 (Hall et al., 1990; Narod et al., 1991). A second locus, BRCA2, has recently been

mapped to chromosome 13q (Wooster et al., 1994) and appears to account for a proportion of early-onset breast cancer roughly equal to BRCA1, but confers a lower risk of ovarian cancer. The remaining susceptibility to early-onset breast cancer is divided between as yet unmapped genes for familial cancer, and rarer germline mutations in genes such as TP53 (Malkin et al., 1990). It has also been suggested that heterozygote carriers for defective forms of the Ataxia-Telangiectasia gene are at higher risk for breast cancer (Swift et al., 1976; Swift et al., 1991). Late-age onset breast cancer is also often familial although the risks in relatives are not as high as those for early-onset breast cancer (Cannon-Albright et al., 1994; Mettlin et al., 1990). However, the percentage of such cases due to genetic susceptibility is unknown.

Breast cancer has long been recognized to be, in part, a familial disease (Anderson, 1972). Numerous investigators have examined the evidence for genetic inheritance and concluded that the data are most consistent with dominant inheritance for a major susceptibility locus or loci (Bishop and Gardner, 1980; Go et al., 1983; Willams and Anderson, 1984; Bishop et al., 1988; Newman et al., 1988; Claus et al., 1991). Recent results demonstrate that at least three loci exist which convey susceptibility to breast cancer as well as other cancers. These loci are the TP53 locus on chromosome 17p (Malkin et al., 1990), a 17q-linked susceptibility locus known as BRCA1 (Hall et al., 1990), and one or more loci responsible for the unmapped residual. Hall et al. (1990) indicated that the inherited breast cancer susceptibility in kindreds with early age onset is linked to chromosome 17q21; although subsequent studies by this group using a more appropriate genetic model partially refuted the limitation to early onset breast cancer (Margaritte et al., 1992).

Most strategies for cloning the 17q-linked breast cancer predisposing gene (BRCA1) require precise genetic localization studies. The simplest model for the functional role of BRCA1 holds that alleles of BRCA1 that predispose to cancer are recessive to wild type alleles; that is, cells that contain at least one wild type BRCA1 allele are not cancerous. However, cells that contain one wild type BRCA1 allele and one predisposing allele may occasionally suffer loss of the wild type allele either by random mutation or by chromosome loss during cell division (nondisjunction). All the progeny of such a mutant cell lack the wild type function of BRCA1 and may develop into tumors. According to this model, predisposing alleles of BRCA1 are recessive, yet susceptibility to cancer is inherited in a dominant fashion: women who possess one predisposing allele (and one wild type allele) risk developing cancer, because their mammary epithelial cells may spontaneously lose the wild type BRCA1 allele. This model applies to a group of cancer susceptibility loci known as tumor suppressors or antioncogenes, a class of genes that includes the retinoblastoma gene and neurofibromatosis gene. By inference this model may also explain the BRCA1 function, as has recently been suggested (Smith et al., 1992).

A second possibility is that BRCA1 predisposing alleles are truly dominant; that is, a wild type allele of BRCA1 cannot overcome the tumor forming role of the predisposing allele. Thus, a cell that carries both wild type and mutant alleles would not necessarily lose the wild type copy of BRCA1 before giving rise to malignant cells. Instead, mammary cells in predisposed individuals would undergo some other stochastic change(s) leading to cancer.

If BRCA1 predisposing alleles are recessive, the BRCA1 gene is expected to be expressed in normal mammary tissue but not functionally expressed in mammary tumors. In

contrast, if BRCA1 predisposing alleles are dominant, the wild type BRCA1 gene may or may not be expressed in normal mammary tissue. However, the predisposing allele will likely be expressed in breast tumor cells.

The 17q linkage of BRCA1 was independently confirmed in three of five kindreds with both breast cancer and ovarian cancer (Narod et al., 1991). These studies claimed to localize the gene within a very large region, 15 centimorgans (cM), or approximately 15 million base pairs, to either side of the linked marker pCMM86 (D17S74). However, attempts to define the region further by genetic studies, using markers surrounding pCMM86, proved unsuccessful. Subsequent studies indicated that the gene was considerably more proximal (Easton et al., 1993) and that the original analysis was flawed (Margaritte et al., 1992). Hall et al., (1992) recently localized the BRCA1 gene to an approximately 8 cM interval (approximately 8 million base pairs) bounded by Mfd15 (D17S250) on the proximal side and the human GIP gene on the distal side. A slightly narrower interval for the BRCA1 locus, based on publicly available data, was agreed upon at the Chromosome 17 workshop in March of 1992 (Fain, 1992). The size of these regions and the uncertainty associated with them has made it exceedingly difficult to design and implement physical mapping and/or cloning strategies for isolating the BRCA1 gene.

Identification of a breast cancer susceptibility locus would permit the early detection of susceptible individuals and greatly increase our ability to understand the initial steps which lead to cancer. As susceptibility loci are often altered during tumor progression, cloning these genes could also be important in the development of better diagnostic and prognostic products, as well as better cancer therapies.

SUMMARY OF THE INVENTION

The present invention relates generally to the field of human genetics. Specifically, the present invention relates to methods and materials used to isolate and detect a human breast cancer predisposing gene (BRCA1), some alleles of which cause susceptibility to cancer, in particular breast and ovarian cancer. More specifically, the present invention relates to germline mutations in the BRCA1 gene and their use in the diagnosis of predisposition to breast and ovarian cancer. The invention further relates to somatic mutations in the BRCA1 gene in human breast cancer and their use in the diagnosis and prognosis of human breast and ovarian cancer. Additionally, the invention relates to somatic mutations in the BRCA1 gene in other human cancers and their use in the diagnosis and prognosis of human cancers. The invention also relates to the therapy of human cancers which have a mutation in the BRCA1 gene, including gene therapy, protein replacement therapy and protein mimetics. The invention further relates to the screening of drugs for cancer therapy. Finally, the invention relates to the screening of the BRCA1 gene for mutations, which are useful for diagnosing the predisposition to breast and ovarian cancer.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 is a diagram showing the order of loci neighboring BRCA1 as determined by the chromosome 17 workshop. FIG. 1 is reproduced from Fain, 1992.

FIG. 2 is a schematic map of YACs which define part of Mfd15-Mfd188 region.

FIG. 3 is a schematic map of STSs, PIs and BACs in the BRCA1 region.

FIG. 4 is a schematic map of human chromosome 17. The pertinent region containing BRCA1 is expanded to indicate

5,753,441

5

the relative positions of two previously identified genes, CA125 and RNU2. BRCA1 spans the marker D17S855.

FIG. 5 shows alignment of the BRCA1 zinc-finger domain with 3 other zinc-finger domains that scored highest in a Smith-Waterman alignment. RPT1 encodes a protein that appears to be a negative regulator of the IL-2 receptor in mouse. RIN 1 encodes a DNA-binding protein that includes a RING-finger motif related to the zinc-finger. RFP1 encodes a putative transcription factor that is the N-terminal domain of the RET oncogene product. The bottom line contains the C3HC4 consensus zinc-finger sequence showing the positions of cysteines and one histidine that form the zinc ion binding pocket.

FIG. 6 is a diagram of BRCA1 mRNA showing the locations of introns and the variants of BRCA1 mRNA produced by alternative splicing. Intron locations are shown by dark triangles and the exons are numbered below the line representing the CDNA. The top cDNA is the composite used to generate the peptide sequence of BRCA1. Alternative forms identified as cDNA clones or hybrid selection clones are shown below.

FIG. 7 shows the tissue expression pattern of BRCA1. The blot was obtained from Clontech and contains RNA from the indicated tissues. Hybridization conditions were as recommended by the manufacturer using a probe consisting of nucleotide positions 3631 to 3930 of BRCA1. Note that both breast and ovary are heterogeneous tissues and the percentage of relevant epithelial cells can be variable. Molecular weight standards are in kilobases.

FIG. 8 is a diagram of the 5' untranslated region plus the beginning of the translated region of BRCA1 showing the locations of introns and the variants of BRCA1 mRNA produced by alternative splicing. Intron locations are shown by broken dashed lines. Six alternate splice forms are shown.

FIG. 9A shows a nonsense mutation in Kindred 2082. P indicates the person originally screened, b and c are haplotype carriers, a, d, e, f, and g do not carry the BRCA1 haplotype. The C to T mutation results in a stop codon and creates a site for the restriction enzyme AvrII. PCR amplification products are cut with this enzyme. The carriers are heterozygous for the site and therefore show three bands. Non-carriers remain uncut.

FIG. 9B shows a mutation and cosegregation analysis in BRCA1 kindreds. Carrier individuals are represented as filled circles and squares in the pedigree diagrams. Frameshift mutation in Kindred 1910. The first three lanes are control, noncarrier samples. Lanes labeled 1-3 contain sequences from carrier individuals. Lane 4 contains DNA from a kindred member who does not carry the BRCA1 mutation. The diamond is used to prevent identification of the kindred. The frameshift resulting from the additional C is apparent in lanes labeled 1, 2, and 3.

FIG. 9C shows a mutation and cosegregation analysis in BRCA1 kindreds. Carrier individuals are represented as filled circles and squares in the pedigree diagrams. Inferred regulatory mutation in Kindred 2035. ASO analysis of carriers and noncarriers of 2 different polymorphisms (PM1 and PM7) which were examined for heterozygosity in the germline and compared to the heterozygosity of lymphocyte mRNA. The top 2 rows of each panel contain PCR products amplified from genomic DNA and the bottom 2 rows contain PCR products amplified from CDNA. "A" and "G" are the two alleles detected by the ASO. The dark spots indicate that a particular allele is present in the sample. The first three lanes of PM7 represent the three genotypes in the general population.

6

FIGS. 10A-10H show genomic sequence of BRCA1. The lower case letters denote intron sequence while the upper case letters denote exon sequence. Indefinite intervals within introns are designated with vvvvvvvvvvvv. Known polymorphic sites are shown as underlined and boldface type.

DETAILED DESCRIPTION OF THE INVENTION

The present invention relates generally to the field of human genetics. Specifically, the present invention relates to methods and materials used to isolate and detect a human breast cancer predisposing gene (BRCA1), some alleles of which cause susceptibility to cancer, in particular breast and ovarian cancer. More specifically, the present invention relates to germline mutations in the BRCA1 gene and their use in the diagnosis of predisposition to breast and ovarian cancer. The invention further relates to somatic mutations in the BRCA1 gene in human breast cancer and their use in the diagnosis and prognosis of human breast and ovarian cancer. Additionally, the invention relates to somatic mutations in the BRCA1 gene in other human cancers and their use in the diagnosis and prognosis of human cancers. The invention also relates to the therapy of human cancers which have a mutation in the BRCA1 gene, including gene therapy, protein replacement therapy and protein mimetics. The invention further relates to the screening of drugs for cancer therapy. Finally, the invention relates to the screening of the BRCA1 gene for mutations, which are useful for diagnosing the predisposition to breast and ovarian cancer.

The present invention provides an isolated polynucleotide comprising all, or a portion of the BRCA1 locus or of a mutated BRCA1 locus, preferably at least eight bases and not more than about 100 kb in length. Such polynucleotides may be antisense polynucleotides. The present invention also provides a recombinant construct comprising such an isolated polynucleotide, for example, a recombinant construct suitable for expression in a transformed host cell.

Also provided by the present invention are methods of detecting a polynucleotide comprising a portion of the BRCA1 locus or its expression product in an analyte. Such methods may further comprise the step of amplifying the portion of the BRCA1 locus, and may further include a step of providing a set of polynucleotides which are primers for amplification of said portion of the BRCA1 locus. The method is useful for either diagnosis of the predisposition to cancer or the diagnosis or prognosis of cancer.

The present invention also provides isolated antibodies, preferably monoclonal antibodies, which specifically bind to an isolated polypeptide comprised of at least five amino acid residues encoded by the BRCA1 locus.

The present invention also provides kits for detecting in an analyte a polynucleotide comprising a portion of the BRCA1 locus, the kits comprising a polynucleotide complementary to the portion of the BRCA1 locus packaged in a suitable container, and instructions for its use.

The present invention further provides methods of preparing a polynucleotide comprising polymerizing nucleotides to yield a sequence comprised of at least eight consecutive nucleotides of the BRCA1 locus; and methods of preparing a polypeptide comprising polymerizing amino acids to yield a sequence comprising at least five amino acids encoded within the BRCA1 locus.

The present invention further provides methods of screening the BRCA1 gene to identify mutations. Such methods may further comprise the step of amplifying a portion of the BRCA1 locus, and may further include a step of providing

5,753,441

7

a set of polynucleotides which are primers for amplification of said portion of the BRCA1 locus. The method is useful for identifying mutations for use in either diagnosis of the predisposition to cancer or the diagnosis or prognosis of cancer.

The present invention further provides methods of screening suspected BRCA1 mutant alleles to identify mutations in the BRCA1 gene.

In addition, the present invention provides methods of screening drugs for cancer therapy to identify suitable drugs for restoring BRCA1 gene product function.

Finally, the present invention provides the means necessary for production of gene-based therapies directed at cancer cells. These therapeutic agents may take the form of polynucleotides comprising all or a portion of the BRCA1 locus placed in appropriate vectors or delivered to target cells in more direct ways such that the function of the BRCA1 protein is reconstituted. Therapeutic agents may also take the form of polypeptides based on either a portion of, or the entire protein sequence of BRCA1. These may functionally replace the activity of BRCA1 in vivo.

It is a discovery of the present invention that the BRCA1 locus which predisposes individuals to breast cancer and ovarian cancer, is a gene encoding a BRCA1 protein, which has been found to have no significant homology with known protein or DNA sequences. This gene is termed BRCA1 herein. It is a discovery of the present invention that mutations in the BRCA1 locus in the germline are indicative of a predisposition to breast cancer and ovarian cancer. Finally, it is a discovery of the present invention that somatic mutations in the BRCA1 locus are also associated with breast cancer, ovarian cancer and other cancers, which represents an indicator of these cancers or of the prognosis of these cancers. The mutational events of the BRCA1 locus can involve deletions, insertions and point mutations within the coding sequence and the non-coding sequence.

Starting from a region on the long arm of human chromosome 17 of the human genome, 17q, which has a size estimated at about 8 million base pairs, a region which contains a genetic locus, BRCA1, which causes susceptibility to cancer, including breast and ovarian cancer, has been identified.

The region containing the BRCA1 locus was identified using a variety of genetic techniques. Genetic mapping techniques initially defined the BRCA1 region in terms of recombination with genetic markers. Based upon studies of large extended families ("kindreds") with multiple cases of breast cancer (and ovarian cancer cases in some kindreds), a chromosomal region has been pinpointed that contains the BRCA1 gene as well as other putative susceptibility alleles in the BRCA1 locus. Two meiotic breakpoints have been discovered on the distal side of the BRCA1 locus which are expressed as recombinants between genetic markers and the disease, and one recombinant on the proximal side of the BRCA1 locus. Thus, a region which contains the BRCA1 locus is physically bounded by these markers.

The use of the genetic markers provided by this invention allowed the identification of clones which cover the region from a human yeast artificial chromosome (YAC) or a human bacterial artificial chromosome (BAC) library. It also allowed for the identification and preparation of more easily manipulated cosmid, P1 and BAC clones from this region and the construction of a from a subset of the clones. These cosmids, P1s, YACs and BACs provide the basis for cloning the BRCA1 locus and provide the basis for developing reagents effective, for example, in the diagnosis and treat-

8

ment of breast and/or ovarian cancer. The BRCA1 gene and other potential susceptibility genes have been isolated from this region. The isolation was done using software trapping (a computational method for identifying sequences likely to contain coding exons, from contiguous or discontinuous genomic DNA sequences), hybrid selection techniques and direct screening, with whole or partial cDNA inserts from cosmids, P1s and BACs, in the region to screen cDNA libraries. These methods were used to obtain sequences of loci expressed in breast and other tissue. These candidate loci were analyzed to identify sequences which confer cancer susceptibility. We have discovered that there are mutations in the coding sequence of the BRCA1 locus in kindreds which are responsible for the 17q-linked cancer susceptibility known as BRCA1. This gene was not known to be in this region. The present invention not only facilitates the early detection of certain cancers, so vital to patient survival, but also permits the detection of susceptible individuals before they develop cancer.

20 Population Resources

Large, well-documented Utah kindreds are especially important in providing good resources for human genetic studies. Each large kindred independently provides the power to detect whether a BRCA1 susceptibility allele is segregating in that family. Recombinants informative for localization and isolation of the BRCA1 locus could be obtained only from kindreds large enough to confirm the presence of a susceptibility allele. Large sibships are especially important for studying breast cancer, since penetrance of the BRCA1 susceptibility allele is reduced both by age and sex, making informative sibships difficult to find. Furthermore, large sibships are essential for constructing haplotypes of deceased individuals by inference from the haplotypes of their close relatives.

While other populations may also provide beneficial information, such studies generally require much greater effort, and the families are usually much smaller and thus less informative. Utah's age-adjusted breast cancer incidence is 20% lower than the average U.S. rate. The lower incidence in Utah is probably due largely to an early age at first pregnancy, increasing the probability that cases found in Utah kindreds carry a genetic predisposition.

Genetic Mapping

Given a set of informative families, genetic markers are essential for linking a disease to a region of a chromosome. Such markers include restriction fragment length polymorphisms (RFLPs) (Botstein et al., 1980), markers with a variable number of tandem repeats (VNTRs) (Jeffreys et al., 1985; Nakamura et al., 1987), and an abundant class of DNA polymorphisms based on short tandem repeats (STRs), especially repeats of CpA (Weber and May, 1989; Litt et al., 1989). To generate a genetic map, one selects potential genetic markers and tests them using DNA extracted from members of the kindreds being studied.

Genetic markers useful in searching for a genetic locus associated with a disease can be selected on an ad hoc basis, by densely covering a specific chromosome, or by detailed analysis of a specific region of a chromosome. A preferred method for selecting genetic markers linked with a disease involves evaluating the degree of informativeness of kindreds to determine the ideal distance between genetic markers of a given degree of polymorphism, then selecting markers from known genetic maps which are ideally spaced for maximal efficiency. Informativeness of kindreds is measured by the probability that the markers will be heterozygous in unrelated individuals. It is also most efficient to use STR markers which are detected by amplification of the

A000274

target nucleic acid sequence using PCR; such markers are highly informative, easy to assay (Weber and May, 1989), and can be assayed simultaneously using multiplexing strategies (Skolnick and Wallace, 1988), greatly reducing the number of experiments required.

Once linkage has been established, one needs to find markers that flank the disease locus, i.e., one or more markers proximal to the disease locus, and one or more markers distal to the disease locus. Where possible, candidate markers can be selected from a known genetic map. Where none is known, new markers can be identified by the STR technique, as shown in the Examples.

Genetic mapping is usually an iterative process. In the present invention, it began by defining flanking genetic markers around the BRCA1 locus, then replacing these flanking markers with other markers that were successively closer to the BRCA1 locus. As an initial step, recombination events, defined by large extended kindreds, helped specifically to localize the BRCA1 locus as either distal or proximal to a specific genetic marker (Goldgar et al., 1994).

The region surrounding BRCA1, until the disclosure of the present invention, was not well mapped and there were few markers. Therefore, short repetitive sequences on cosmids subcloned from YACs, which had been physically mapped, were analyzed in order to develop new genetic markers. Using this approach, one marker of the present invention, 42D6, was discovered which replaced pCMM86 as the distal flanking marker for the BRCA1 region. Since 42D6 is approximately 14 cM from pCMM86, the BRCA1 region was thus reduced by approximately 14 centimorgans (Easton et al., 1993). The present invention thus began by finding a much more closely linked distal flanking marker of the BRCA1 region. BRCA1 was then discovered to be distal to the genetic marker Mfd15. Therefore, BRCA1 was shown to be in a region of 6 to 10 million bases bounded by Mfd15 and 42D6. Marker Mfd191 was subsequently discovered to be distal to Mfd15 and proximal to BRCA1. Thus, Mfd15 was replaced with Mfd191 as the closest proximal genetic marker. Similarly, it was discovered that genetic marker Mfd188 could replace genetic marker 42D6, narrowing the region containing the BRCA1 locus to approximately 1.5 million bases. Then the marker Mfd191 was replaced with tdj 1474 as the proximal marker and Mfd188 was replaced with U5R as the distal marker, further narrowing the BRCA1 region to a small enough region to allow isolation and characterization of the BRCA1 locus (see FIG. 3), using techniques known in the art and described herein.

Physical Mapping

Three distinct methods were employed to physically map the region. The first was the use of yeast artificial chromosomes (YACs) to clone the region which is flanked by tdj 1474 and U5R. The second was the creation of a set of P1, BAC and cosmid clones which cover the region containing the BRCA1 locus.

Yeast Artificial Chromosomes (YACs).

Once a sufficiently small region containing the BRCA1 locus was identified, physical isolation of the DNA in the region proceeded by identifying a set of overlapping YACs which covers the region. Useful YACs can be isolated from known libraries, such as the St. Louis and CEPH YAC libraries, which are widely distributed and contain approximately 50,000 YACs each. The YACs isolated were from these publicly accessible libraries and can be obtained from a number of sources including the Michigan Genome Center. Clearly, others who had access to these YACs, without the disclosure of the present invention, would not have known the value of the specific YACs we selected since they

would not have known which YACs were within, and which YACs outside of, the smallest region containing the BRCA1 locus.

Cosmid, P1 and BAC Clones.

In the present invention, it is advantageous to proceed by obtaining cosmid, P1, and BAC clones to cover this region. The smaller size of these inserts, compared to YAC inserts, makes them more useful as specific hybridization probes. Furthermore, having the cloned DNA in bacterial cells, rather than in yeast cells, greatly increases the ease with which the DNA of interest can be manipulated, and improves the signal-to-noise ratio of hybridization assays. For cosmid subclones of YACs, the DNA is partially digested with the restriction enzyme Sau3A and cloned into the BamHI site of the pWE15 cosmid vector (Stratagene, cat. #1251201). The cosmids containing human sequences are screened by hybridization with human repetitive DNA (e.g., Gibco/BRL, Hfuman C₀t-1 DNA, cat. 5279SA), and then fingerprinted by a variety of techniques, as detailed in the Examples.

P1 and BAC clones are obtained by screening libraries constructed from the total human genome with specific sequence tagged sites (STSs) derived from the YACs, cosmids or P1s and BACs, isolated as described herein.

These P1, BAC and cosmid clones can be compared by interspersed repetitive sequence (IRS) PCR and/or restriction enzyme digests followed by gel electrophoresis and comparison of the resulting DNA fragments ("fingerprints") (Maniatis et al., 1982). The clones can also be characterized by the presence of STSs. The fingerprints are used to define an overlapping contiguous set of clones which covers the region but is not excessively redundant, referred to herein as a "minimum tiling path". Such a minimum tiling path forms the basis for subsequent experiments to identify cDNAs which may originate from the BRCA1 locus.

Coverage of the Gap with P1 and BAC Clones.

To cover any gaps in the BRCA1 contig between the identified cosmids with genomic clones, clones in P1 and BAC vectors which contain inserts of genomic DNA roughly twice as large as cosmids for P1s and still greater for BACs (Stenberg, 1990; Stenberg et al., 1990; Pierce et al., 1992; Shizuya et al., 1992) were used. P1 clones were isolated by Genome Sciences using PCR primers provided by us for screening. BACs were provided by hybridization techniques in Dr. Mel Simon's laboratory. The strategy of using P1 clones also permitted the covering of the genomic region with an independent set of clones not derived from YACs. This guards against the possibility of other deletions in YACs that have not been detected. These new sequences derived from the P1 clones provide the material for further screening for candidate genes, as described below.

Gene Isolation

There are many techniques for testing genomic clones for the presence of sequences likely to be candidates for the coding sequence of a locus one is attempting to isolate, including but not limited to:

- a. zoo blots
- b. identifying HTF islands
- c. exon trapping
- d. hybridizing cDNA to cosmids or YACs.
- e. screening cDNA libraries.
- (a) Zoo blots.

The first technique is to hybridize cosmids to Southern blots to identify DNA sequences which are evolutionarily conserved, and which therefore give positive hybridization signals with DNA from species of varying degrees of

relationship to humans (such as monkey, cow, chicken, pig, mouse and rat). Southern blots containing such DNA from a variety of species are commercially available (Clontech, Cat. 7753-1).

(b) Identifying HTF islands.

The second technique involves finding regions rich in the nucleotides C and G, which often occur near or within coding sequences. Such sequences are called HTF (HpaI tiny fragment) or CpG islands, as restriction enzymes specific for sites which contain CpG dimers cut frequently in these regions (Lindsay et al, 1987).

(c) Exon trapping.

The third technique is exon trapping, a method that identifies sequences in genomic DNA which contain splice junctions and therefore are likely to comprise coding sequences of genes. Exon amplification (Buckler et al., 1991) is used to select and amplify exons from DNA clones described above. Exon amplification is based on the selection of RNA sequences which are flanked by functional 5' and/or 3' splice sites. The products of the exon amplification are used to screen the breast cDNA libraries to identify a manageable number of candidate genes for further study. Exon trapping can also be performed on small segments of sequenced DNA using computer programs or by software trapping.

(d) Hybridizing cDNA to Cosmids, PIs, BACs or YACs.

The fourth technique is a modification of the selective enrichment technique which utilizes hybridization of cDNA to cosmids, PIs, BACs or YACs and permits transcribed sequences to be identified in, and recovered from cloned genomic DNA (Kandpal et al., 1990). The selective enrichment technique, as modified for the present purpose, involves binding DNA from the region of BRCA1 present in a YAC to a column matrix and selecting cDNAs from the relevant libraries which hybridize with the bound DNA, followed by amplification and purification of the bound DNA, resulting in a great enrichment for cDNAs in the region represented by the cloned genomic DNA.

(e) Identification of cDNAs.

The fifth technique is to identify cDNAs that correspond to the BRCA1 locus. Hybridization probes containing putative coding sequences, selected using any of the above techniques, are used to screen various libraries, including breast tissue CDNA libraries, ovarian CDNA libraries, and any other necessary libraries.

Another variation on the theme of direct selection of CDNA was also used to find candidate genes for BRCA1 (Lovett et al, 1991; Futreal, 1993). This method uses cosmid, P1 or BAC DNA as the probe. The probe DNA is digested with a blunt cutting restriction enzyme such as HaeIII. Double stranded adapters are then ligated onto the DNA and serve as binding sites for primers in subsequent PCR amplification reactions using biotinylated primers. Target CDNA is generated from mRNA derived from tissue samples, e.g., breast tissue, by synthesis of either random primed or oligo(dT) primed first strand followed by second strand synthesis. The CDNA ends are rendered blunt and ligated onto double-stranded adapters. These adapters serve as amplification sites for PCR. The target and probe sequences are denatured and mixed with human C₀t-1 DNA to block repetitive sequences. Solution hybridization is carried out to high C₀t-1/2 values to ensure hybridization of rare target cDNA molecules. The annealed material is then captured on avidin beads, washed at high stringency and the retained cDNAs are eluted and amplified by PCR. The selected cDNA is subjected to further rounds of enrichment before cloning into a plasmid vector for analysis.

Testing the cDNA for Candidacy

Proof that the cDNA is the BRCA1 locus is obtained by finding sequences in DNA extracted from affected kindred members which create abnormal BRCA1 gene products or abnormal levels of BRCA1 gene product. Such BRCA1 susceptibility alleles will co-segregate with the disease in large kindreds. They will also be present at a much higher frequency in non-kindred individuals with breast and ovarian cancer than in individuals in the general population. Finally, since tumors often mutate somatically at loci which are in other instances mutated in the germline, we expect to see normal germline BRCA1 alleles mutated into sequences which are identical or similar to BRCA1 susceptibility alleles in DNA extracted from tumor tissue. Whether one is comparing BRCA1 sequences from tumor tissue to BRCA1 alleles from the germline of the same individuals, or one is comparing germline BRCA1 alleles from cancer cases to those from unaffected individuals, the key is to find mutations which are serious enough to cause obvious disruption to the normal function of the gene product. These mutations can take a number of forms. The most severe forms would be frame shift mutations or large deletions which would cause the gene to code for an abnormal protein or one which would significantly alter protein expression. Less severe disruptive mutations would include small in-frame deletions and nonconservative base pair substitutions which would have a significant effect on the protein produced, such as changes to or from a cysteine residue, from a basic to an acidic amino acid or vice versa, from a hydrophobic to hydrophilic amino acid or vice versa, or other mutations which would affect secondary, tertiary or quaternary protein structure. Silent mutations or those resulting in conservative amino acid substitutions would not generally be expected to disrupt protein function.

According to the diagnostic and prognostic method of the present invention, alteration of the wild-type BRCA1 locus is detected. In addition, the method can be performed by detecting the wild-type BRCA1 locus and confirming the lack of a predisposition to cancer at the BRCA1 locus. "Alteration of a wild-type gene" encompasses all forms of mutations including deletions, insertions and point mutations in the coding and noncoding regions. Deletions may be of the entire gene or of only a portion of the gene. Point mutations may result in stop codons, frameshift mutations or amino acid substitutions. Somatic mutations are those which occur only in certain tissues, e.g., in the tumor tissue, and are not inherited in the germline. Germline mutations can be found in any of a body's tissues and are inherited. If only a single allele is somatically mutated, an early neoplastic state is indicated. However, if both alleles are somatically mutated, then a late neoplastic state is indicated. The finding of BRCA1 mutations thus provides both diagnostic and prognostic information. A BRCA1 allele which is not deleted (e.g., found on the sister chromosome to a chromosome carrying a BRCA1 deletion) can be screened for other mutations, such as insertions, small deletions, and point mutations. It is believed that many mutations found in tumor tissues will be those leading to decreased expression of the BRCA1 gene product. However, mutations leading to non-functional gene products would also lead to a cancerous state. Point mutational events may occur in regulatory regions, such as in the promoter of the gene, leading to loss or diminution of expression of the mRNA. Point mutations may also abolish proper RNA processing, leading to loss of expression of the BRCA1 gene product, or to a decrease in mRNA stability or translation efficiency.

Useful diagnostic techniques include, but are not limited to fluorescent in situ hybridization (FISH), direct DNA

5,753,441

13

sequencing, PFGE analysis, Southern blot analysis, single stranded conformation analysis (SSCA), RNase protection assay, allele-specific oligonucleotide (ASO), dot blot analysis and PCR-SSCP, as discussed in detail further below.

Predisposition to cancers, such as breast and ovarian cancer, and the other cancers identified herein, can be ascertained by testing any tissue of a human for mutations of the BRCA1 gene. For example, a person who has inherited a germline BRCA1 mutation would be prone to develop cancers. This can be determined by testing DNA from any tissue of the person's body. Most simply, blood can be drawn and DNA extracted from the cells of the blood. In addition, prenatal diagnosis can be accomplished by testing fetal cells, placental cells or amniotic cells for mutations of the BRCA1 gene. Alteration of a wild-type BRCA1 allele, whether, for example, by point mutation or deletion, can be detected by any of the means discussed herein.

There are several methods that can be used to detect DNA sequence variation. Direct DNA sequencing, either manual sequencing or automated fluorescent sequencing can detect sequence variation. For a gene as large as BRCA1, manual sequencing is very labor-intensive, but under optimal conditions, mutations in the coding sequence of a gene are rarely missed. Another approach is the single-stranded conformation polymorphism assay (SSCA) (Orita et al., 1989). This method does not detect all sequence changes, especially if the DNA fragment size is greater than 200 bp, but can be optimized to detect most DNA sequence variation. The reduced detection sensitivity is a disadvantage, but the increased throughput possible with SSCA makes it an attractive, viable alternative to direct sequencing for mutation detection on a research basis. The fragments which have shifted mobility on SSCA gels are then sequenced to determine the exact nature of the DNA sequence variation. Other approaches based on the detection of mismatches between the two complementary DNA strands include clamped denaturing gel electrophoresis (CDGE) (Sheffield et al., 1991), heteroduplex analysis (HA) (White et al., 1992) and chemical mismatch cleavage (CMC) (Grompe et al., 1989). None of the methods described above will detect large deletions, duplications or insertions, nor will they detect a regulatory mutation which affects transcription or translation of the protein. Other methods which might detect these classes of mutations such as a protein truncation assay or the asymmetric assay, detect only specific types of mutations and would not detect missense mutations. A review of currently available methods of detecting DNA sequence variation can be found in a recent review by Grompe (1993). Once a mutation is known, an allele specific detection approach such as allele specific oligonucleotide (ASO) hybridization can be utilized to rapidly screen large numbers of other samples for that same mutation.

In order to detect the alteration of the wild-type BRCA1 gene in a tissue, it is helpful to isolate the tissue free from surrounding normal tissues. Means for enriching tissue preparation for tumor cells are known in the art. For example, the tissue may be isolated from paraffin or cryostat sections. Cancer cells may also be separated from normal cells by flow cytometry. These techniques, as well as other techniques for separating tumor cells from normal cells, are well known in the art. If the tumor tissue is highly contaminated with normal cells, detection of mutations is more difficult.

A rapid preliminary analysis to detect polymorphisms in DNA sequences can be performed by looking at a series of Southern blots of DNA cut with one or more restriction enzymes, preferably with a large number of restriction

14

enzymes. Each blot contains a series of normal individuals and a series of cancer cases, tumors, or both. Southern blots displaying hybridizing fragments (differing in length from control DNA when probed with sequences near or including the BRCA1 locus) indicate a possible mutation. If restriction enzymes which produce very large restriction fragments are used, then pulsed field gel electrophoresis (PFGE) is employed.

Detection of point mutations may be accomplished by molecular cloning of the BRCA1 allele(s) and sequencing the allele(s) using techniques well known in the art. Alternatively, the gene sequences can be amplified directly from a genomic DNA preparation from the tumor tissue, using known techniques. The DNA sequence of the amplified sequences can then be determined.

There are six well known methods for a more complete, yet still indirect, test for confirming the presence of a susceptibility allele: 1) single stranded conformation analysis (SSCA) (Orita et al., 1989); 2) denaturing gradient gel electrophoresis (DGGE) (Wartell et al., 1990; Sheffield et al., 1989); 3) RNase protection assays (Finkelstein et al., 1990; Kinszler et al., 1991); 4) allele-specific oligonucleotides (ASOs) (Conner et al., 1983); 5) the use of proteins which recognize nucleotide mismatches, such as the *E. coli* mutS protein (Modrich, 1991); and 6) allele-specific PCR (Rano & Kidd, 1989). For allele-specific PCR, primers are used which hybridize at their 3' ends to a particular BRCA1 mutation. If the particular BRCA1 mutation is not present, an amplification product is not observed. Amplification Refractory Mutation System (ARMS) can also be used, as disclosed in European Patent Application Publication No. 0332435 and in Newton et al., 1989. Insertions and deletions of genes can also be detected by cloning, sequencing and amplification. In addition, restriction fragment length polymorphism (RFLP) probes for the gene or surrounding marker genes can be used to score alteration of an allele or an insertion in a polymorphic fragment. Such a method is particularly useful for screening relatives of an affected individual for the presence of the BRCA1 mutation found in that individual. Other techniques for detecting insertions and deletions as known in the art can be used.

In the first three methods (SSCA, DGGE and RNase protection assay), a new electrophoretic band appears. SSCA detects a band which migrates differentially because the sequence change causes a difference in single-strand, intramolecular base pairing. RNase protection involves cleavage of the mutant polynucleotide into two or more smaller fragments. DGGE detects differences in migration rates of mutant sequences compared to wild-type sequences, using a denaturing gradient gel. In an allele-specific oligonucleotide assay, an oligonucleotide is designed which detects a specific sequence, and the assay is performed by detecting the presence or absence of a hybridization signal. In the mutS assay, the protein binds only to sequences that contain a nucleotide mismatch in a heteroduplex between mutant and wild-type sequences.

Mismatches, according to the present invention, are hybridized nucleic acid duplexes in which the two strands are not 100% complementary. Lack of total homology may be due to deletions, insertions, inversions or substitutions. Mismatch detection can be used to detect point mutations in the gene or in its mRNA product. While these techniques are less sensitive than sequencing, they are simpler to perform on a large number of tumor samples. An example of a mismatch cleavage technique is the RNase protection method. In the practice of the present invention, the method involves the use of a labeled riboprobe which is comple-

mentary to the human wild-type BRCA1 gene coding sequence. The riboprobe and either mRNA or DNA isolated from the tumor tissue are annealed (hybridized) together and subsequently digested with the enzyme RNase A which is able to detect some mismatches in a duplex RNA structure. If a mismatch is detected by RNase A, it cleaves at the site of the mismatch. Thus, when the annealed RNA preparation is separated on an electrophoretic gel matrix, if a mismatch has been detected and cleaved by RNase A, an RNA product will be seen which is smaller than the full length duplex RNA for the riboprobe and the mRNA or DNA. The riboprobe need not be the full length of the BRCA1 mRNA or gene but can be a segment of either. If the riboprobe comprises only a segment of the BRCA1 mRNA or gene, it will be desirable to use a number of these probes to screen the whole mRNA sequence for mismatches.

In similar fashion, DNA probes can be used to detect mismatches, through enzymatic or chemical cleavage. See, e.g., Cotton et al., 1988; Shenk et al., 1975; Novack et al., 1986. Alternatively, mismatches can be detected by shifts in the electrophoretic mobility of mismatched duplexes relative to matched duplexes. See, e.g., Cariello, 1988. With either riboprobes or DNA probes, the cellular mRNA or DNA which might contain a mutation can be amplified using PCR (see below) before hybridization. Changes in DNA of the BRCA1 gene can also be detected using Southern hybridization, especially if the changes are gross rearrangements, such as deletions and insertions.

DNA sequences of the BRCA1 gene which have been amplified by screened using allele-specific probes. These probes are nucleic acid oligomers, each of which contains a region of the BRCA1 gene sequence harboring a known mutation. For example, one oligomer may be about 30 nucleotides in length, corresponding to a portion of the BRCA1 gene sequence. By use of a battery of such allele-specific probes, PCR amplification products can be screened to identify the presence of a previously identified mutation in the BRCA1 gene. Hybridization of allele-specific probes with amplified BRCA1 sequences can be performed, for example, on a nylon filter. Hybridization to a particular probe under stringent hybridization conditions indicates the presence of the same mutation in the tumor tissue as in the allele-specific probe.

The most definitive test for mutations in a candidate locus is to directly compare genomic BRCA1 sequences from cancer patients with those from a control population. Alternatively, one could sequence messenger RNA after amplification, e.g., by PCR, thereby eliminating the necessity of determining the exon structure of the candidate gene.

Mutations from cancer patients falling outside the coding region of BRCA1 can be detected by examining the non-coding regions, such as introns and regulatory sequences near or within the BRCA1 gene. An early indication that mutations in noncoding regions are important may come from Northern blot experiments that reveal messenger RNA molecules of abnormal size or abundance in cancer patients as compared to control individuals.

Alteration of BRCA1 mRNA expression can be detected by any techniques known in the art. These include Northern blot analysis, PCR amplification and RNase protection. Diminished mRNA expression indicates an alteration of the wild-type BRCA1 gene. Alteration of wild-type BRCA1 genes can also be detected by screening for alteration of wild-type BRCA1 protein. For example, monoclonal antibodies immunoreactive with BRCA1 can be used to screen a tissue. Lack of cognate antigen would indicate a BRCA1 mutation. Antibodies specific for products of mutant alleles

could also be used to detect mutant BRCA1 gene product. Such immunological assays can be done in any convenient formats known in the art. These include Western blots, immunohistochemical assays and ELISA assays. Any means for detecting an altered BRCA1 protein can be used to detect alteration of wild-type BRCA1 genes. Functional assays, such as protein binding determinations, can be used. In addition, assays can be used which detect BRCA1 biochemical function. Finding a mutant BRCA1 gene product indicates alteration of a wild-type BRCA1 gene.

Mutant BRCA1 genes or gene products can also be detected in other human body samples, such as serum, stool, urine and sputum. The same techniques discussed above for detection of mutant BRCA1 genes or gene products in tissues can be applied to other body samples. Cancer cells are sloughed off from tumors and appear in such body samples. In addition, the BRCA1 gene product itself may be secreted into the extracellular space and found in these body samples even in the absence of cancer cells. By screening such body samples, a simple early diagnosis can be achieved for many types of cancers. In addition, the progress of chemotherapy or radiotherapy can be monitored more easily by testing such body samples for mutant BRCA1 genes or gene products.

The methods of diagnosis of the present invention are applicable to any tumor in which BRCA1 has a role in tumorigenesis. The diagnostic method of the present invention is useful for clinicians, so they can decide upon an appropriate course of treatment.

The primer pairs of the present invention are useful for determination of the nucleotide sequence of a particular BRCA1 allele using PCR. The pairs of single-stranded DNA primers can be annealed to sequences within or surrounding the BRCA1 gene on chromosome 17q21 in order to prime amplifying DNA synthesis of the BRCA1 gene itself. A complete set of these primers allows synthesis of all of the nucleotides of the BRCA1 gene coding sequences, i.e., the exons. The set of primers preferably allows synthesis of both intron and exon sequences. Allele-specific primers can also be used. Such primers anneal only to particular BRCA1 mutant alleles, and thus will only amplify a product in the presence of the mutant allele as a template.

In order to facilitate subsequent cloning of amplified sequences, primers may have restriction enzyme site sequences appended to their 5' ends. Thus, all nucleotides of the primers are derived from BRCA1 sequences or sequences adjacent to BRCA1, except for the few nucleotides necessary to form a restriction enzyme site. Such enzymes and sites are well known in the art. The primers themselves can be synthesized using techniques which are well known in the art. Generally, the primers can be made using oligonucleotide synthesizing machines which are commercially available. Given the sequence of the BRCA1 open reading frame shown in SEQ ID NO:1, design of particular primers is well within the skill of the art.

The nucleic acid probes provided by the present invention are useful for a number of purposes. They can be used in Southern hybridization to genomic DNA and in the RNase protection method for detecting point mutations already discussed above. The probes can be used to detect PCR amplification products. They may also be used to detect mismatches with the BRCA1 gene or mRNA using other techniques.

It has been discovered that individuals with the wild-type BRCA1 gene do not have cancer which results from the BRCA1 allele. However, mutations which interfere with the function of the BRCA1 protein are involved in the patho-

5,753,441

17

genesis of cancer. Thus, the presence of an altered (or a mutant) BRCA1 gene which produces a protein having a loss of function, or altered function, directly correlates to an increased risk of cancer. In order to detect a BRCA1 gene mutation, a biological sample is prepared and analyzed for a difference between the sequence of the BRCA1 allele being analyzed and the sequence of the wild-type BRCA1 allele. Mutant BRCA1 alleles can be initially identified by any of the techniques described above. The mutant alleles are then sequenced to identify the specific mutation of the particular mutant allele. Alternatively, mutant BRCA1 alleles can be initially identified by identifying mutant (altered) BRCA1 proteins, using conventional techniques. The mutant alleles are then sequenced to identify the specific mutation for each allele. The mutations, especially those which lead to an altered function of the BRCA1 protein, are then used for the diagnostic and prognostic methods of the present invention.

Definitions

The present invention employs the following definitions: "Amplification of Polynucleotides" utilizes methods such as the polymerase chain reaction (PCR), ligation amplification (or ligase chain reaction, LCR) and amplification methods based on the use of Q-beta replicase. These methods are well known and widely practiced in the art. See, e.g., U.S. Pat. Nos. 4,683,195 and 4,683,202 and Innis et al., 1990 (for PCR); and Wu et al., 1989a (for LCR). Reagents and hardware for conducting PCR are commercially available. Primers useful to amplify sequences from the BRCA1 region are preferably complementary to, and hybridize specifically to sequences in the BRCA1 region or in regions that flank a target region therein. BRCA1 sequences generated by amplification may be sequenced directly. Alternatively, but less desirably, the amplified sequence(s) may be cloned prior to sequence analysis. A method for the direct cloning and sequence analysis of enzymatically amplified genomic segments has been described by Scharf, 1986.

"Analyte polynucleotide" and "analyte strand" refer to a single- or double-stranded polynucleotide which is suspected of containing a target sequence, and which may be present in a variety of types of samples, including biological samples.

"Antibodies." The present invention also provides polyclonal and/or monoclonal antibodies and fragments thereof, and immunologic binding equivalents thereof, which are capable of specifically binding to the BRCA1 polypeptides and fragments thereof or to polynucleotide sequences from the BRCA1 region, particularly from the BRCA1 locus or a portion thereof. The term "antibody" is used both to refer to a homogeneous molecular entity, or a mixture such as a serum product made up of a plurality of different molecular entities. Polypeptides may be prepared synthetically in a peptide synthesizer and coupled to a carrier molecule (e.g., keyhole limpet hemocyanin) and injected over several months into rabbits. Rabbit sera is tested for immunoreactivity to the BRCA1 polypeptide or fragment. Monoclonal antibodies may be made by injecting mice with the protein polypeptides, fusion proteins or fragments thereof. Monoclonal antibodies will be screened by ELISA and tested for specific immunoreactivity with BRCA1 polypeptide or fragments thereof. See, Harlow & Lane, 1988. These antibodies will be useful in assays as well as pharmaceuticals.

Once a sufficient quantity of desired polypeptide has been obtained, it may be used for various purposes. A typical use is the production of antibodies specific for binding. These antibodies may be either polyclonal or monoclonal, and may be produced by in vitro or in vivo techniques well known in

18

the art. For production of polyclonal antibodies, an appropriate target immune system, typically mouse or rabbit, is selected. Substantially purified antigen is presented to the immune system in a fashion determined by methods appropriate for the animal and by other parameters well known to immunologists. Typical sites for injection are in footpads, intramuscularly, intraperitoneally, or intradermally. Of course, other species may be substituted for mouse or rabbit. Polyclonal antibodies are then purified using techniques known in the art, adjusted for the desired specificity.

An immunological response is usually assayed with an immunoassay. Normally, such immunoassays involve some purification of a source of antigen, for example, that produced by the same cells and in the same fashion as the antigen. A variety of immunoassay methods are well known in the art. See, e.g., Harlow & Lane, 1988, or Goding, 1986.

Monoclonal antibodies with affinities of 10^{-8} M⁻¹ or preferably 10^{-9} to 10^{-10} to M⁻¹ or stronger will typically be made by standard procedures as described, e.g., in Harlow & Lane, 1988 or Goding, 1986. Briefly, appropriate animals will be selected and the desired immunization protocol followed. After the appropriate period of time, the spleens of such animals are excised and individual spleen cells fused, typically, to immortalized myeloma cells under appropriate selection conditions. Thereafter, the cells are clonally separated and the supernatants of each clone tested for their production of an appropriate antibody specific for the desired region of the antigen.

Other suitable techniques involve in vitro exposure of lymphocytes to the antigenic polypeptides, or alternatively, to selection of libraries of antibodies in phage or similar vectors. See Huse et al., 1989. The polypeptides and antibodies of the present invention may be used with or without modification. Frequently, polypeptides and antibodies will be labeled by joining, either covalently or non-covalently, a substance which provides for a detectable signal. A wide variety of labels and conjugation techniques are known and are reported extensively in both the scientific and patent literature. Suitable labels include radionuclides, enzymes, substrates, cofactors, inhibitors, fluorescent agents, chemiluminescent agents, magnetic particles and the like. Patents teaching the use of such labels include U.S. Pat. Nos. 3,817,837; 3,850,752; 3,939,350; 3,996,345; 4,277,437; 4,275,149 and 4,366,241. Also, recombinant immunoglobulins may be produced (see U.S. Pat. No. 4,816,567).

"Binding partner" refers to a molecule capable of binding a ligand molecule with high specificity, as for example, an antigen and an antigen-specific antibody or an enzyme and its inhibitor. In general, the specific binding partners must bind with sufficient affinity to immobilize the analyte copy/complementary strand duplex (in the case of polynucleotide hybridization) under the isolation conditions. Specific binding partners are known in the art and include, for example, biotin and avidin or streptavidin, IgG and protein A, the numerous, known receptor-ligand couples, and complementary polynucleotide strands. In the case of complementary polynucleotide binding partners, the partners are normally at least about 15 bases in length, and may be at least 40 bases in length. The polynucleotides may be composed of DNA, RNA, or synthetic nucleotide analogs.

A "biological sample" refers to a sample of tissue or fluid suspected of containing an analyte polynucleotide or polypeptide from an individual including, but not limited to, e.g., plasma, serum, spinal fluid, lymph fluid, the external sections of the skin, respiratory, intestinal, and genitourinary tracts, tears, saliva, blood cells, tumors, organs, tissue and samples of in vitro cell culture constituents.

A000279

5,753,441

19

As used herein, the terms "diagnosing" or "prognosing," as used in the context of neoplasia, are used to indicate 1) the classification of lesions as neoplasia, 2) the determination of the severity of the neoplasia, or 3) the monitoring of the disease progression, prior to, during and after treatment.

"Encode". A polynucleotide is said to "encode" a polypeptide if, in its native state or when manipulated by methods well known to those skilled in the art, it can be transcribed and/or translated to produce the mRNA for and/or the polypeptide or a fragment thereof. The anti-sense strand is the complement of such a nucleic acid, and the encoding sequence can be deduced therefrom.

"Isolated" or "substantially pure". An "isolated" or "substantially pure" nucleic acid (e.g., an RNA, DNA or a mixed polymer) is one which is substantially separated from other cellular components which naturally accompany a native human sequence or protein, e.g., ribosomes, polymerases, many other human genome sequences and proteins. The term embraces a nucleic acid sequence or protein which has been removed from its naturally occurring environment, and includes recombinant or cloned DNA isolates and chemically synthesized analogs or analogs biologically synthesized by heterologous systems.

"BRCA1 Allele" refers to normal alleles of the BRCA1 locus as well as alleles carrying variations that predispose individuals to develop cancer of many sites including, for example, breast, ovarian, colorectal and prostate cancer. Such predisposing alleles are also called "BRCA1 susceptibility alleles".

"BRCA1 Locus," "BRCA1 Gene," "BRCA1 Nucleic Acids" or "BRCA1 Polynucleotide" each refer to polynucleotides, all of which are in the BRCA1 region, that are likely to be expressed in normal tissue, certain alleles of which predispose an individual to develop breast, ovarian, colorectal and prostate cancers. Mutations at the BRCA1 locus may be involved in the initiation and/or progression of other types of tumors. The locus is indicated in part by mutations that predispose individuals to develop cancer. These mutations fall within the BRCA1 region described infra. The BRCA1 locus is intended to coding sequences, intervening sequences and regulatory elements controlling transcription and/or translation. The BRCA1 locus is intended to include all allelic variations of the DNA sequence.

These terms, when applied to a nucleic acid, refer to a nucleic acid which encodes a BRCA1 polypeptide, fragment, homolog or variant, including, e.g., protein fusions or deletions. The nucleic acids of the present invention will possess a sequence which is either derived from, or substantially similar to a natural BRCA1-encoding gene or one having substantial homology with a natural BRCA1-encoding gene or a portion thereof. The coding sequence for a BRCA1 polypeptide is shown in SEQ ID NO:1, with the amino acid sequence shown in SEQ ID NO:2.

The polynucleotide compositions of this invention include RNA, cDNA, genomic DNA, synthetic forms, and mixed polymers, both sense and antisense strands, and may be chemically or biochemically modified or may contain non-natural or derivatized nucleotide bases, as will be readily appreciated by those skilled in the art. Such modifications include, for example, labels, methylation, substitution of one or more of the naturally occurring nucleotides with an analog, internucleotide modifications such as uncharged linkages (e.g., methyl phosphonates, phosphotriesters, phosphoamidates, carbamates, etc.), charged linkages (e.g., phosphorothioates, phosphorodithioates, etc.), pendent moieties (e.g.,

20

polypeptides), intercalators (e.g., acridine, psoralen, etc.), chelators, alkylators, and modified linkages (e.g., alpha anomeric nucleic acids, etc.). Also included are synthetic molecules that mimic polynucleotides in their ability to bind to a designated sequence via hydrogen bonding and other chemical interactions. Such molecules are known in the art and include, for example, those in which peptide linkages substitute for phosphate linkages in the backbone of the molecule.

The present invention provides recombinant nucleic acids comprising all or part of the BRCA1 region. The recombinant construct may be capable of replicating autonomously in a host cell. Alternatively, the recombinant construct may become integrated into the chromosomal DNA of the host cell. Such a recombinant polynucleotide comprises a polynucleotide of genomic, cDNA, semi-synthetic, or synthetic origin which, by virtue of its origin or manipulation, 1) is not associated with all or a portion of a polynucleotide with which it is associated in nature; 2) is linked to a polynucleotide other than that to which it is linked in nature; or 3) does not occur in nature.

Therefore, recombinant nucleic acids comprising sequences otherwise not naturally occurring are provided by this invention. Although the wild-type sequence may be employed, it will often be altered, e.g., by deletion, substitution or insertion.

cDNA or genomic libraries of various types may be screened as natural sources of the nucleic acids of the present invention, or such nucleic acids may be provided by amplification of sequences resident in genomic DNA or other natural sources, e.g., by PCR. The choice of cDNA libraries normally corresponds to a tissue source which is abundant in mRNA for the desired proteins. Phage libraries are normally preferred, but other types of libraries may be used. Clones of a library are spread onto plates, transferred to a substrate for screening, denatured and probed for the presence of desired sequences.

The DNA sequences used in this invention will usually comprise at least about five codons (15 nucleotides), more usually at least about 7-15 codons, and most preferably, at least about 35 codons. One or more introns may also be present. This number of nucleotides is usually about the minimal length required for a successful probe that would hybridize specifically with a BRCA1-encoding sequence.

Techniques for nucleic acid manipulation are described generally, for example, in Sambrook et al., 1989 or Ausubel et al., 1992. Reagents useful in applying such techniques, such as restriction enzymes and the like, are widely known in the art and commercially available from such vendors as New England BioLabs, Boehringer Mannheim, Amersham, Promega Biotec, U. S. Biochemicals, New England Nuclear, and a number of other sources. The recombinant nucleic acid sequences used to produce fusion proteins of the present invention may be derived from natural or synthetic sequences. Many natural gene sequences are obtainable from various cDNA or from genomic libraries using appropriate probes. See, GenBank, National Institutes of Health.

"BRCA1 Region" refers to a portion of human chromosome 17q21 bounded by the markers tdj1474 and U5R. This region contains the BRCA1 locus, including the BRCA1 gene.

As used herein, the terms "BRCA1 locus," "BRCA1 allele" and "BRCA1 region" all refer to the double-stranded DNA comprising the locus, allele, or region, as well as either of the single-stranded DNAs comprising the locus, allele or region.

As used herein, a "portion" of the BRCA1 locus or region or allele is defined as having a minimal size of at least about

A000280

5,753,441

21

eight nucleotides, or preferably about 15 nucleotides, or more preferably at least about 25 nucleotides, and may have a minimal size of at least about 40 nucleotides.

"BRCA1 protein" or "BRCA1 polypeptide" refer to a protein or polypeptide encoded by the BRCA1 locus, variants or fragments thereof. The term "polypeptide" refers to a polymer of amino acids and its equivalent and does not refer to a specific length of the product; thus, peptides, oligopeptides and proteins are included within the definition of a polypeptide. This term also does not refer to, or exclude modifications of the polypeptide, for example, glycosylations, acetylations, phosphorylations, and the like. Included within the definition are, for example, polypeptides containing one or more analogs of an amino acid (including, for example, unnatural amino acids, etc.), polypeptides with substituted linkages as well as other modifications known in the art, both naturally and non-naturally occurring. Ordinarily, such polypeptides will be at least about 50% homologous to the native BRCA1 sequence, preferably in excess of about 90%, and more preferably at least about 95% homologous. Also included are proteins encoded by DNA which hybridize under high or low stringency conditions, to BRCA1-encoding nucleic acids and closely related polypeptides or proteins retrieved by antisera to the BRCA1 protein (s).

The length of polypeptide sequences compared for homology will generally be at least about 16 amino acids, usually at least about 20 residues, more usually at least about 24 residues, typically at least about 28 residues, and preferably more than about 35 residues.

"Operably linked" refers to a juxtaposition wherein the components so described are in a relationship permitting them to function in their intended manner. For instance, a promoter is operably linked to a coding sequence if the promoter affects its transcription or expression.

"Probes". Polynucleotide polymorphisms associated with BRCA1 alleles which predispose to certain cancers or are associated with most cancers are detected by hybridization with a polynucleotide probe which forms a stable hybrid with that of the target sequence, under stringent to moderately stringent hybridization and wash conditions. If it is expected that the probes will be perfectly complementary to the target sequence, stringent conditions will be used. Hybridization stringency may be lessened if some mismatching is expected, for example, if variants are expected with the result that the probe will not be completely complementary. Conditions are chosen which rule out nonspecific/adventitious bindings, that is, which minimize noise. Since such indications identify neutral DNA polymorphisms as well as mutations, these indications need further analysis to demonstrate detection of a BRCA1 susceptibility allele.

Probes for BRCA1 alleles may be derived from the sequences of the BRCA1 region or its cDNAs. The probes may be of any suitable length, which span all or a portion of the BRCA1 region, and which allow specific hybridization to the BRCA1 region. If the target sequence contains a sequence identical to that of the probe, the probes may be short, e.g., in the range of about 8-30 base pairs, since the hybrid will be relatively stable under even stringent conditions. If some degree of mismatch is expected with the probe, i.e., if it is suspected that the probe will hybridize to a variant region, a longer probe may be employed which hybridizes to the target sequence with the requisite specificity.

The probes will include an isolated polynucleotide attached to a label or reporter molecule and may be used to isolate other polynucleotide sequences, having sequence

22

similarity by standard methods. For techniques for preparing and labeling probes see, e.g., Sambrook et al., 1989 or Ausubel et al., 1992. Other similar polynucleotides may be selected by using homologous polynucleotides. Alternatively, polynucleotides encoding these or similar polypeptides may be synthesized or selected by use of the redundancy in the genetic code. Various codon substitutions may be introduced, e.g., by silent changes (thereby producing various restriction sites) or to optimize expression for a particular system. Mutations may be introduced to modify the properties of the polypeptide, perhaps to change ligand-binding affinities, interchain affinities, or the polypeptide degradation or turnover rate.

Probes comprising synthetic oligonucleotides or other polynucleotides of the present invention may be derived from naturally occurring or recombinant single- or double-stranded polynucleotides, or be chemically synthesized. Probes may also be labeled by nick translation, Klenow fill-in reaction, or other methods known in the art.

Portions of the polynucleotide sequence having at least about eight nucleotides, usually at least about 15 nucleotides, and fewer than about 6 kb, usually fewer than about 1.0 kb, from a polynucleotide sequence encoding BRCA1 are preferred as probes. The probes may also be used to determine whether mRNA encoding BRCA1 is present in a cell or tissue.

"Protein modifications or fragments" are provided by the present invention for BRCA1 polypeptides or fragments thereof which are substantially homologous to primary structural sequence but which include, e.g., in vivo or in vitro chemical and biochemical modifications or which incorporate unusual amino acids. Such modifications include, for example, acetylation, carboxylation, phosphorylation, glycosylation, ubiquitination, labeling, e.g., with radionuclides, and various enzymatic modifications, as will be readily appreciated by those well skilled in the art. A variety of methods for labeling polypeptides and of substituents or labels useful for such purposes are well known in the art, and include radioactive isotopes such as ^{32}P , ligands which bind to labeled antigens (e.g., antibodies), fluorophores, chemiluminescent agents, enzymes, and antigens which can serve as specific binding pair members for a labeled ligand. The choice of label depends on the sensitivity required, ease of conjugation with the primer, stability requirements, and available instrumentation. Methods of labeling polypeptides are well known in the art. See, e.g., Sambrook et al., 1989 or Ausubel et al., 1992.

Besides substantially full-length polypeptides, the present invention provides for biologically active fragments of the polypeptides. Significant biological activities include ligand-binding, immunological activity and other biological activities characteristic of BRCA1 polypeptides. Immunological activities include both immunogenic function in a target immune system, as well as sharing of immunological epitopes for binding, serving as either a competitor or substitute antigen for an epitope of the BRCA1 protein. As used herein, "epitope" refers to an antigenic determinant of a polypeptide. An epitope could comprise three amino acids in a spatial conformation which is unique to the epitope. Generally, an epitope consists of at least five such amino acids, and more usually consists of at least 8-10 such amino acids. Methods of determining the spatial conformation of such amino acids are known in the art.

For immunological purposes, tandem-repeat polypeptide segments may be used as immunogens, thereby producing highly antigenic proteins. Alternatively, such polypeptides

5,753,441

23

will serve as highly efficient competitors for specific binding. Production of antibodies specific for BRCA1 polypeptides or fragments thereof is described below.

The present invention also provides for fusion polypeptides, comprising BRCA1 polypeptides and fragments. Homologous polypeptides may be fusions between two or more BRCA1 polypeptide sequences or between the sequences of BRCA1 and a related protein. Likewise, heterologous fusions may be constructed which would exhibit a combination of properties or activities of the derivative proteins. For example, ligand-binding or other domains may be "swapped" between different new fusion polypeptides or fragments. Such homologous or heterologous fusion polypeptides may display, for example, altered strength or specificity of binding. Fusion partners include immunoglobulins, bacterial β -galactosidase, trpE, protein A, β -lactamase, alpha amylase, alcohol dehydrogenase and yeast alpha mating factor. See, e.g., Godowski et al., 1988.

Fusion proteins will typically be made by either recombinant nucleic acid methods, as described below, or may be chemically synthesized. Techniques for the synthesis of polypeptides are described, for example, in Merrifield, 1963.

"Protein purification" refers to various methods for the isolation of the BRCA1 polypeptides from other biological material, such as from cells transformed with recombinant nucleic acids encoding BRCA1, and are well known in the art. For example, such polypeptides may be purified by immunoaffinity chromatography employing, e.g., the antibodies provided by the present invention. Various methods of protein purification are well known in the art, and include those described in Deutscher, 1990 and Scopes, 1982.

The terms "isolated", "substantially pure", and "substantially homogeneous" are used interchangeably to describe a protein or polypeptide which has been separated from components which accompany it in its natural state. A monomeric protein is substantially pure when at least about 60 to 75% of a sample exhibits a single polypeptide sequence. A substantially pure protein will typically comprise about 60 to 90% W/W of a protein sample, more usually about 95%, and preferably will be over about 99% pure. Protein purity or homogeneity may be indicated by a number of means well known in the art, such as polyacrylamide gel electrophoresis of a protein sample, followed by visualizing a single polypeptide band upon staining the gel. For certain purposes, higher resolution may be provided by using HPLC or other means well known in the art which are utilized for purification.

A BRCA1 protein is substantially free of naturally associated components when it is separated from the native contaminants which accompany it in its natural state. Thus, a polypeptide which is chemically synthesized or synthesized in a cellular system different from the cell from which it naturally originates will be substantially free from its naturally associated components. A protein may also be rendered substantially free of naturally associated components by isolation, using protein purification techniques well known in the art.

A polypeptide produced as an expression product of an isolated and manipulated genetic sequence is an "isolated polypeptide," as used herein, even if expressed in a homologous cell type. Synthetically made forms or molecules expressed by heterologous cells are inherently isolated molecules.

"Recombinant nucleic acid" is a nucleic acid which is not naturally occurring, or which is made by the artificial combination of two otherwise separated segments of sequence. This artificial combination is often accomplished

24

by either chemical synthesis means, or by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques. Such is usually done to replace a codon with a redundant codon encoding the same or a conservative amino acid, while typically introducing or removing a sequence recognition site. Alternatively, it is performed to join together nucleic acid segments of desired functions to generate a desired combination of functions.

"Regulatory sequences" refers to those sequences normally within 100 kb of the coding region of a locus, but they may also be more distant from the coding region, which affect the expression of the gene (including transcription of the gene, and translation, splicing, stability or the like of the messenger RNA).

"Substantial homology or similarity". A nucleic acid or fragment thereof is "substantially homologous" ("or substantially similar") to another if, when optimally aligned (with appropriate nucleotide insertions or deletions) with the other nucleic acid (or its complementary strand), there is nucleotide sequence identity in at least about 60% of the nucleotide bases, usually at least about 70%, more usually at least about 80%, preferably at least about 90%, and more preferably at least about 95-98% of the nucleotide bases.

Alternatively, substantial homology or (similarity) exists when a nucleic acid or fragment thereof will hybridize to another nucleic acid (or a complementary strand thereof) under selective hybridization conditions, to a strand, or to its complement. Selectivity of hybridization exists when hybridization which is substantially more selective than total lack of specificity occurs. Typically, selective hybridization will occur when there is at least about 55% homology over a stretch of at least about 14 nucleotides, preferably at least about 65%, more preferably at least about 75%, and most preferably at least about 90%. See, Kanehisa, 1984. The length of homology comparison, as described, may be over longer stretches, and in certain embodiments will often be over a stretch of at least about nine nucleotides, usually at least about 20 nucleotides, more usually at least about 24 nucleotides, typically at least about 28 nucleotides, more typically at least about 32 nucleotides, and preferably at least about 36 or more nucleotides.

Nucleic acid hybridization will be affected by such conditions as salt concentration, temperature, or organic solvents, in addition to the base composition, length of the complementary strands, and the number of nucleotide base mismatches between the hybridizing nucleic acids, as will be readily appreciated by those skilled in the art. Stringent temperature conditions will generally include temperatures in excess of 30° C., typically in excess of 37° C., and preferably in excess of 45° C. Stringent salt conditions will ordinarily be less than 1000 mM, typically less than 500 mM, and preferably less than 200 mM. However, the combination of parameters is much more important than the measure of any single parameter. See, e.g., Wetmur & Davidson, 1968.

Probe sequences may also hybridize specifically to duplex DNA under certain conditions to form triplex or other higher order DNA complexes. The preparation of such probes and suitable hybridization conditions are well known in the art.

The terms "substantial homology" or "substantial identity", when referring to polypeptides, indicate that the polypeptide or protein in question exhibits at least about 30% identity with an entire naturally-occurring protein or a portion thereof, usually at least about 70% identity, and preferably at least about 95% identity.

"Substantially similar function" refers to the function of a modified nucleic acid or a modified protein, with reference

A000282

5,753,441

25

to the wild-type BRCA1 nucleic acid or wild-type BRCA1 polypeptide. The modified polypeptide will be substantially homologous to the wild-type BRCA1 polypeptide and will have substantially the same function. The modified polypeptide may have an altered amino acid sequence and/or may contain modified amino acids. In addition to the similarity of function, the modified polypeptide may have other useful properties, such as a longer half-life. The similarity of function (activity) of the modified polypeptide may be substantially the same as the activity of the wild-type BRCA1 polypeptide. Alternatively, the similarity of function (activity) of the modified polypeptide may be higher than the activity of the wild-type BRCA1 polypeptide. The modified polypeptide is synthesized using conventional techniques, or is encoded by a modified nucleic acid and produced using conventional techniques. The modified nucleic acid is prepared by conventional techniques. A nucleic acid with a function substantially similar to the wild-type BRCA1 gene function produces the modified protein described above.

Homology, for polypeptides, is typically measured using sequence analysis software. See, e.g., the Sequence Analysis Software Package of the Genetics Computer Group, University of Wisconsin Biotechnology Center, 910 University Avenue, Madison, Wis. 53705. Protein analysis software matches similar sequences using measure of homology assigned to various substitutions, deletions and other modifications. Conservative substitutions typically include substitutions within the following groups: glycine, alanine; valine, isoleucine, leucine; aspartic acid, glutamic acid; asparagine, glutamine; serine, threonine; lysine, arginine; and phenylalanine, tyrosine.

A polypeptide "fragment," "portion" or "segment" is a stretch of amino acid residues of at least about five to seven contiguous amino acids, often at least about seven to nine contiguous amino acids, typically at least about nine to 13 contiguous amino acids and, most preferably, at least about 20 to 30 or more contiguous amino acids.

The polypeptides of the present invention, if soluble, may be coupled to a solid-phase support, e.g., nitrocellulose, nylon, column packing materials (e.g., Sepharose beads), magnetic beads, glass wool, plastic, metal, polymer gels, cells, or other substrates. Such supports may take the form, for example, of beads, wells, dipsticks, or membranes.

"Target region" refers to a region of the nucleic acid which is amplified and/or detected. The term "target sequence" refers to a sequence with which a probe or primer will form a stable hybrid under desired conditions.

The practice of the present invention employs, unless otherwise indicated, conventional techniques of chemistry, molecular biology, microbiology, recombinant DNA, genetics, and immunology. See, e.g., Maniatis et al., 1982; Sambrook et al., 1989; Ausubel et al., 1992; Glover, 1985; Anand, 1992; Guthrie & Fink, 1991. A general discussion of techniques and materials for human gene mapping, including mapping of human chromosome 17q, is provided, e.g., in White and Lalouel, 1988.

Preparation of recombinant or chemically synthesized nucleic acids; vectors, transformation, host cells

Large amounts of the polynucleotides of the present invention may be produced by replication in a suitable host cell. Natural or synthetic polynucleotide fragments coding for a desired fragment will be incorporated into recombinant polynucleotide constructs, usually DNA constructs, capable of introduction into and replication in a prokaryotic or eukaryotic cell. Usually the polynucleotide constructs will be suitable for replication in a unicellular host, such as yeast

26

or bacteria, but may also be intended for introduction to (with and without integration within the genome) cultured mammalian or plant or other eukaryotic cell lines. The purification of nucleic acids produced by the methods of the present invention is described, e.g., in Sambrook et al., 1989 or Ausubel et al., 1992.

The polynucleotides of the present invention may also be produced by chemical synthesis, e.g., by the phosphoramidite method described by Beaucage & Carruthers, 1981 or the triester method according to Matteucci and Caruthers, 1981, and may be performed on commercial, automated oligonucleotide synthesizers. A double-stranded fragment may be obtained from the single-stranded product of chemical synthesis either by synthesizing the complementary strand and annealing the strands together under appropriate conditions or by adding the complementary strand using DNA polymerase with an appropriate primer sequence.

Polynucleotide constructs prepared for introduction into a prokaryotic or eukaryotic host may comprise a replication system recognized by the host, including the intended polynucleotide fragment encoding the desired polypeptide, and will preferably also include transcription and translational initiation regulatory sequences operably linked to the polypeptide encoding segment. Expression vectors may include, for example, an origin of replication or autonomously replicating sequence (ARS) and expression control sequences, a promoter, an enhancer and necessary processing information sites, such as ribosome-binding sites, RNA splice sites, polyadenylation sites, transcriptional terminator sequences, and mRNA stabilizing sequences. Secretion signals may also be included where appropriate, whether from a native BRCA1 protein or from other receptors or from secreted polypeptides of the same or related species, which allow the protein to cross and/or lodge in cell membranes, and thus attain its functional topology, or be secreted from the cell. Such vectors may be prepared by means of standard recombinant techniques well known in the art and discussed, for example, in Sambrook et al., 1989 or Ausubel et al., 1992.

An appropriate promoter and other necessary vector sequences will be selected so as to be functional in the host, and may include, when appropriate, those naturally associated with BRCA1 genes. Examples of workable combinations of cell lines and expression vectors are described in Sambrook et al., 1989 or Ausubel et al., 1992; see also, e.g., Metzger et al., 1988. Many useful vectors are known in the art and may be obtained from such vendors as Stratagene, New England Biolabs, Promega Biotech, and others. Promoters such as the trp, lac and phage promoters, tRNA promoters and glycolytic enzyme promoters may be used in prokaryotic hosts. Useful yeast promoters include promoter regions for metallothionein, 3-phosphoglycerate kinase or other glycolytic enzymes such as enolase or glyceraldehyde-3-phosphate dehydrogenase, enzymes responsible for maltose and galactose utilization, and others. Vectors and promoters suitable for use in yeast expression are further described in Hitzeman et al., EP 73,675A. Appropriate non-native mammalian promoters might include the early and late promoters from SV40 (Fiers et al., 1978) or promoters derived from murine Moloney leukemia virus, mouse tumor virus, avian sarcoma viruses, adenovirus II, bovine papilloma virus or polyoma. In addition, the construct may be joined to an amplifiable gene (e.g., DHFR) so that multiple copies of the gene may be made. For appropriate enhancer and other expression control sequences, see also *Enhancers and Eukaryotic Gene Expression*, Cold Spring Harbor Press, Cold Spring Harbor, N.Y. (1983).

While such expression vectors may replicate autonomously, they may also replicate by being inserted into the genome of the host cell, by methods well known in the art.

A000283

5,753,441

27

Expression and cloning vectors will likely contain a selectable marker, a gene encoding a protein necessary for survival or growth of a host cell transformed with the vector. The presence of this gene ensures growth of only those host cells which express the inserts. Typical selection genes encode proteins that a) confer resistance to antibiotics or other toxic substances, e.g. ampicillin, neomycin, methotrexate, etc.; b) complement auxotrophic deficiencies, or c) supply critical nutrients not available from complex media, e.g., the gene encoding D-alanine racemase for *Bacilli*. The choice of the proper selectable marker will depend on the host cell, and appropriate markers for different hosts are well known in the art.

The vectors containing the nucleic acids of interest can be transcribed *in vitro*, and the resulting RNA introduced into the host cell by well-known methods, e.g., by injection (see, Kubo et al., 1988), or the vectors can be introduced directly into host cells by methods well known in the art, which vary depending on the type of cellular host, including electroporation; transfection employing calcium chloride, rubidium chloride, calcium phosphate, DEAE-dextran, or other substances; microprojectile bombardment; lipofection; infection (where the vector is an infectious agent, such as a retroviral genome); and other methods. See generally, Sambrook et al., 1989 and Ausubel et al., 1992. The introduction of the polynucleotides into the host cell by any method known in the art, including, *inter alia*, those described above, will be referred to herein as "transformation." The cells into which have been introduced nucleic acids described above are meant to also include the progeny of such cells.

Large quantities of the nucleic acids and polypeptides of the present invention may be prepared by expressing the BRCA1 nucleic acids or portions thereof in vectors or other expression vehicles in compatible prokaryotic or eukaryotic host cells. The most commonly used prokaryotic hosts are strains of *Escherichia coli*, although other prokaryotes, such as *Bacillus subtilis* or *Pseudomonas* may also be used.

Mammalian or other eukaryotic host cells, such as those of yeast, filamentous fungi, plant, insect, or amphibian or avian species, may also be useful for production of the proteins of the present invention. Propagation of mammalian cells in culture is per se well known. See, Jakoby and Pastan, 1979. Examples of commonly used mammalian host cell lines are VERO and HeLa cells, Chinese hamster ovary (CHO) cells, and WI38, BHK, and COS cell lines, although it will be appreciated by the skilled practitioner that other cell lines may be appropriate, e.g., to provide higher expression, desirable glycosylation patterns, or other features.

Clones are selected by using markers depending on the mode of the vector construction. The marker may be on the same or a different DNA molecule, preferably the same DNA molecule. In prokaryotic hosts, the transformant may be selected, e.g., by resistance to ampicillin, tetracycline or other antibiotics. Production of a particular product based on temperature sensitivity may also serve as an appropriate marker.

Prokaryotic or eukaryotic cells transformed with the polynucleotides of the present invention will be useful not only for the production of the nucleic acids and polypeptides of the present invention, but also, for example, in studying the characteristics of BRCA1 polypeptides.

Antisense polynucleotide sequences are useful in preventing or diminishing the expression of the BRCA1 locus, as will be appreciated by those skilled in the art. For example, polynucleotide vectors containing all or a portion of the BRCA1 locus or other sequences from the BRCA1 region

28

(particularly those flanking the BRCA1 locus) may be placed under the control of a promoter in an antisense orientation and introduced into a cell. Expression of such an antisense construct within a cell will interfere with BRCA1 transcription and/or translation and/or replication.

The probes and primers based on the BRCA1 gene sequences disclosed herein are used to identify homologous BRCA1 gene sequences and proteins in other species. These BRCA1 gene sequences and proteins are used in the diagnostic/prognostic, therapeutic and drug screening methods described herein for the species from which they have been isolated.

Methods of Use: Nucleic Acid Diagnosis and Diagnostic Kits

In order to detect the presence of a BRCA1 allele predisposing an individual to cancer, a biological sample such as blood is prepared and analyzed for the presence or absence of susceptibility alleles of BRCA1. In order to detect the presence of neoplasia, the progression toward malignancy of a precursor lesion, or as a prognostic indicator, a biological sample of the lesion is prepared and analyzed for the presence or absence of mutant alleles of BRCA1. Results of these tests and interpretive information are returned to the health care provider for communication to the tested individual. Such diagnoses may be performed by diagnostic laboratories, or, alternatively, diagnostic kits are manufactured and sold to health care providers or to private individuals for self-diagnosis.

Initially, the screening method involves amplification of the relevant BRCA1 sequences. In another preferred embodiment of the invention, the screening method involves a non-PCR based strategy. Such screening methods include two-step label amplification methodologies that are well known in the art. Both PCR and non-PCR based screening strategies can detect target sequences with a high level of sensitivity.

The most popular method used today is target amplification. Here, the target nucleic acid sequence is amplified with polymerases. One particularly preferred method using polymerase-driven amplification is the polymerase chain reaction (PCR). The polymerase chain reaction and other polymerase-driven amplification assays can achieve over a million-fold increase in copy number through the use of polymerase-driven amplification cycles. Once amplified, the resulting nucleic acid can be sequenced or used as a substrate for DNA probes.

When the probes are used to detect the presence of the target sequences (for example, in screening for cancer susceptibility), the biological sample to be analyzed, such as blood or serum, may be treated, if desired, to extract the nucleic acids. The sample nucleic acid may be prepared in various ways to facilitate detection of the target sequence; e.g. denaturation, restriction digestion, electrophoresis or dot blotting. The targeted region of the analyte nucleic acid usually must be at least partially single-stranded to form hybrids with the targeting sequence of the probe. If the sequence is naturally single-stranded, denaturation will not be required. However, if the sequence is double-stranded, the sequence will probably need to be denatured. Denaturation can be carried out by various techniques known in the art.

Analyte nucleic acid and probe are incubated under conditions which promote stable hybrid formation of the target sequence in the probe with the putative targeted sequence in the analyte. The region of the probes which is used to bind to the analyte can be made completely complementary to the targeted region of human chromosome 17q.

5,753,441

29

Therefore, high stringency conditions are desirable in order to prevent false positives. However, conditions of high stringency are used only if the probes are complementary to regions of the chromosome which are unique in the genome. The stringency of hybridization is determined by a number of factors during hybridization and during the washing procedure, including temperature, ionic strength, base composition, probe length, and concentration of formamide. These factors are outlined in, for example, Maniatis et al., 1982 and Sambrook et al., 1989. Under certain circumstances, the formation of higher order hybrids, such as triplexes, quadruplexes, etc., may be desired to provide the means of detecting target sequences.

Detection, if any, of the resulting hybrid is usually accomplished by the use of labeled probes. Alternatively, the probe may be unlabeled, but may be detectable by specific binding with a ligand which is labeled, either directly or indirectly. Suitable labels, and methods for labeling probes and ligands are known in the art, and include, for example, radioactive labels which may be incorporated by known methods (e.g., nick translation, random priming or kinasin), biotin, fluorescent groups, chemiluminescent groups (e.g., dioxetanes, particularly triggered dioxetanes), enzymes, antibodies and the like. Variations of this basic scheme are known in the art, and include those variations that facilitate separation of the hybrids to be detected from extraneous materials and/or that amplify the signal from the labeled moiety. A number of these variations are reviewed in, e.g., Matthews & Kricka, 1988; Landegren et al., 1988; Mittlin, 1989; U.S. Patent No. 4,868,105, and in EPO Publication No. 225,807.

As noted above, non-PCR based screening assays are also contemplated in this invention. An exemplary non-PCR based procedure is provided in Example 11. This procedure hybridizes a nucleic acid probe (or an analog such as a methyl phosphonate backbone replacing the normal phosphodiester), to the low level DNA target. This probe may have an enzyme covalently linked to the probe, such that the covalent linkage does not interfere with the specificity of the hybridization. This enzyme-probe-conjugate-target nucleic acid complex can then be isolated away from the free probe enzyme conjugate and a substrate is added for enzyme detection. Enzymatic activity is observed as a change in color development or luminescent output resulting in a 10^3 – 10^6 increase in sensitivity. For an example relating to the preparation of oligodeoxynucleotide-alkaline phosphatase conjugates and their use as hybridization probes see Jablonski et al., 1986.

Two-step label amplification methodologies are known in the art. These assays work on the principle that a small ligand (such as digoxigenin, biotin, or the like) is attached to a nucleic acid probe capable of specifically binding BRCA1. Exemplary probes are provided in Table 9 of this patent application and additionally include the nucleic acid probe corresponding to nucleotide positions 3631 to 3930 of SEQ ID NO:1. Allele specific probes are also contemplated within the scope of this example and exemplary allele specific probes include probes encompassing the predisposing mutations summarized in Tables 11 and 12 of this patent application.

In one example, the small ligand attached to the nucleic acid probe is specifically recognized by an antibody-enzyme conjugate. In one embodiment of this example, digoxigenin is attached to the nucleic acid probe. Hybridization is detected by an antibody-alkaline phosphatase conjugate which turns over a chemiluminescent substrate. For methods for labeling nucleic acid probes according to this embodiment see Martin et al., 1990. In a second example, the small

30

ligand is recognized by a second ligand-enzyme conjugate that is capable of specifically complexing to the first ligand. A well known embodiment of this example is the biotin-avidin type of interactions. For methods for labeling nucleic acid probes and their use in biotin-avidin based assays see Rigby et al., 1977 and Nguyen et al., 1992.

It is also contemplated within the scope of this invention that the nucleic acid probe assays of this invention will employ a cocktail of nucleic acid probes capable of detecting BRCA1. Thus, in one example to detect the presence of BRCA1 in a cell sample, more than one probe complementary to BRCA1 is employed and in particular the number of different probes is alternatively 2, 3, or 5 different nucleic acid probe sequences. In another example, to detect the presence of mutations in the BRCA1 gene sequence in a patient, more than one probe complementary to BRCA1 is employed where the cocktail includes probes capable of binding to the allele-specific mutations identified in populations of patients with alterations in BRCA1. In this embodiment, any number of probes can be used, and will preferably include probes corresponding to the major gene mutations identified as predisposing an individual to breast cancer. Some candidate probes contemplated within the scope of the invention include probes that include the allele-specific mutations identified in Tables 11 and 12 and those that have the BRCA1 regions corresponding to SEQ ID NO:1 both 5' and 3' to the mutation site.

Methods of Use: Peptide Diagnosis and Diagnostic Kits

The neoplastic condition of lesions can also be detected on the basis of the alteration of wild-type BRCA1 polypeptide. Such alterations can be determined by sequence analysis in accordance with conventional techniques. More preferably, antibodies (polyclonal or monoclonal) are used to detect differences in, or the absence of BRCA1 peptides. The antibodies may be prepared as discussed above under the heading "Antibodies" and as further shown in Examples 12 and 13. Other techniques for raising and purifying antibodies are well known in the art and any such techniques may be chosen to achieve the preparations claimed in this invention. In a preferred embodiment of the invention, antibodies will immunoprecipitate BRCA1 proteins from solution as well as react with BRCA1 protein on Western or immunoblots of polyacrylamide gels. In another preferred embodiment, antibodies will detect BRCA1 proteins in paraffin or frozen tissue sections, using immunocytochemical techniques.

Preferred embodiments relating to methods for detecting BRCA1 or its mutations include enzyme linked immunosorbent assays (ELISA), radioimmunoassays (RIA), immunoradiometric assays (IRMA) and immunoenzymatic assays (IEMA), including sandwich assays using monoclonal and/or polyclonal antibodies. Exemplary sandwich assays are described by David et al. in U.S. Pat. Nos. 4,376,110 and 4,486,530, hereby incorporated by reference, and exemplified in Example 14.

Methods of Use: Drug Screening

This invention is particularly useful for screening compounds by using the BRCA1 polypeptide or binding fragment thereof in any of a variety of drug screening techniques.

The BRCA1 polypeptide or fragment employed in such a test may either be free in solution, affixed to a solid support, or borne on a cell surface. One method of drug screening utilizes eucaryotic or procaryotic host cells which are stably transformed with recombinant polynucleotides expressing the polypeptide or fragment, preferably in competitive binding assays. Such cells, either in viable or fixed form, can be

A000285

5,753,441

31

used for standard binding assays. One may measure, for example, for the formation of complexes between a BRCA1 polypeptide or fragment and the agent being tested, or examine the degree to which the formation of a complex between a BRCA1 polypeptide or fragment and a known ligand is interfered with by the agent being tested.

Thus, the present invention provides methods of screening for drugs comprising contacting such an agent with a BRCA1 polypeptide or fragment thereof and assaying (i) for the presence of a complex between the agent and the BRCA1 polypeptide or fragment, or (ii) for the presence of a complex between the BRCA1 polypeptide or fragment and a ligand, by methods well known in the art. In such competitive binding assays the BRCA1 polypeptide or fragment is typically labeled. Free BRCA1 polypeptide or fragment is separated from that present in a protein:protein complex, and the amount of free (i.e., uncomplexed) label is a measure of the binding of the agent being tested to BRCA1 or its interference with BRCA1:ligand binding, respectively.

Another technique for drug screening provides high throughput screening for compounds having suitable binding affinity to the BRCA1 polypeptides and is described in detail in Geysen, PCT published application WO 84/03564, published on Sep. 13, 1984. Briefly stated, large numbers of different small peptide test compounds are synthesized on a solid substrate, such as plastic pins or some other surface. The peptide test compounds are reacted with BRCA1 polypeptide and washed. Bound BRCA1 polypeptide is then detected by methods well known in the art.

Purified BRCA1 can be coated directly onto plates for use in the aforementioned drug screening techniques. However, non-neutralizing antibodies to the polypeptide can be used to capture antibodies to immobilize the BRCA1 polypeptide on the solid phase.

This invention also contemplates the use of competitive drug screening assays in which neutralizing antibodies capable of specifically binding the BRCA1 polypeptide compete with a test compound for binding to the BRCA1 polypeptide or fragments thereof. In this manner, the antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants of the BRCA1 polypeptide.

A further technique for drug screening involves the use of host eukaryotic cell lines or cells (such as described above) which have a nonfunctional BRCA1 gene. These host cell lines or cells are defective at the BRCA1 polypeptide level. The host cell lines or cells are grown in the presence of drug compound. The rate of growth of the host cells is measured to determine if the compound is capable of regulating the growth of BRCA1 defective cells.

Methods of Use: Rational Drug Design

The goal of rational drug design is to produce structural analogs of biologically active polypeptides of interest or of small molecules with which they interact (e.g., agonists, antagonists, inhibitors) in order to fashion drugs which are, for example, more active or stable forms of the polypeptide, or which, e.g., enhance or interfere with the function of a polypeptide in vivo. See, e.g., Hodgson, 1991. In one approach, one first determines the three-dimensional structure of a protein of interest (e.g., BRCA1 polypeptide) or, for example, of the BRCA1-receptor or ligand complex, by x-ray crystallography, by computer modeling or most typically, by a combination of approaches. Less often, useful information regarding the structure of a polypeptide may be gained by modeling based on the structure of homologous proteins. An example of rational drug design is the development of HIV protease inhibitors (Erickson et al., 1990). In

32

addition, peptides (e.g., BRCA1 polypeptide) are analyzed by an alanine scan (Wells, 1991). In this technique, an amino acid residue is replaced by Ala, and its effect on the peptide's activity is determined. Each of the amino acid residues of the peptide is analyzed in this manner to determine the important regions of the peptide.

It is also possible to isolate a target-specific antibody, selected by a functional assay, and then to solve its crystal structure. In principle, this approach yields a pharmacore upon which subsequent drug design can be based. It is possible to bypass protein crystallography altogether by generating anti-idio-typic antibodies (anti-ids) to a functional, pharmacologically active antibody. As a mirror image of a mirror image, the binding site of the anti-ids would be expected to be an analog of the original receptor. The anti-id could then be used to identify and isolate peptides from banks of chemically or biologically produced banks of peptides. Selected peptides would then act as the pharmacore.

Thus, one may design drugs which have, e.g., improved BRCA1 polypeptide activity or stability or which act as inhibitors, agonists, antagonists, etc. of BRCA1 polypeptide activity. By virtue of the availability of cloned BRCA1 sequences, sufficient amounts of the BRCA1 polypeptide may be made available to perform such analytical studies as x-ray crystallography. In addition, the knowledge of the BRCA1 protein sequence provided herein will guide those employing computer modeling techniques in place of, or in addition to x-ray crystallography.

Methods of Use: Gene Therapy

According to the present invention, a method is also provided of supplying wild-type BRCA1 function to a cell which carries mutant BRCA1 alleles. Supplying such a function should suppress neoplastic growth of the recipient cells. The wild-type BRCA1 gene or a part of the gene may be introduced into the cell in a vector such that the gene remains extrachromosomal. In such a situation, the gene will be expressed by the cell from the extrachromosomal location. If a gene fragment is introduced and expressed in a cell carrying a mutant BRCA1 allele, the gene fragment should encode a part of the BRCA1 protein which is required for non-neoplastic growth of the cell. More preferred is the situation where the wild-type BRCA1 gene or a part thereof is introduced into the mutant cell in such a way that it recombines with the endogenous mutant BRCA1 gene present in the cell. Such recombination requires a double recombination event which results in the correction of the BRCA1 gene mutation. Vectors for introduction of genes both for recombination and for extrachromosomal maintenance are known in the art, and any suitable vector may be used. Methods for introducing DNA into cells such as electroporation, calcium phosphate co-precipitation and viral transduction are known in the art, and the choice of method is within the competence of the routineer. Cells transformed with the wild-type BRCA1 gene can be used as model systems to study cancer remission and drug treatments which promote such remission.

As generally discussed above, the BRCA1 gene or fragment, where applicable, may be employed in gene therapy methods in order to increase the amount of the expression products of such genes in cancer cells. Such gene therapy is particularly appropriate for use in both cancerous and pre-cancerous cells, in which the level of BRCA1 polypeptide is absent or diminished compared to normal cells. It may also be useful to increase the level of expression of a given BRCA1 gene even in those tumor cells in which the mutant gene is expressed at a "normal" level, but the gene product is not fully functional.

A000286

Gene therapy would be carried out according to generally accepted methods, for example, as described by Friedman, 1991. Cells from a patient's tumor would be first analyzed by the diagnostic methods described above, to ascertain the production of BRCA1 polypeptide in the tumor cells. A virus or plasmid vector (see further details below), containing a copy of the BRCA1 gene linked to expression control elements and capable of replicating inside the tumor cells, is prepared. Suitable vectors are known, such as disclosed in U.S. Pat. No. 5,252,479 and PCT published application WO 93/07282. The vector is then injected into the patient, either locally at the site of the tumor or systemically (in order to reach any tumor cells that may have metastasized to other sites). If the transfected gene is not permanently incorporated into the genome of each of the targeted tumor cells, the treatment may have to be repeated periodically.

Gene transfer systems known in the art may be useful in the practice of the gene therapy methods of the present invention. These include viral and nonviral transfer methods. A number of viruses have been used as gene transfer vectors, including papovaviruses, e.g., SV40 (Madzak et al., 1992), adenovirus (Berkner, 1992; Berkner et al., 1988; Gorziglia and Kapikian, 1992; Quantin et al., 1992; Rosenfeld et al., 1992; Wilkinson et al., 1992; Stratford-Perricaudet et al., 1990), vaccinia virus (Moss, 1992), adeno-associated virus (Muzyczka, 1992; Ohi et al., 1990), herpesviruses including HSV and EBV (Margolske, 1992; Johnson et al., 1992; Fink et al., 1992; Breakfield and Geller, 1987; Freese et al., 1990), and retroviruses of avian (Brandyopadhyay and Temin, 1984; Petropoulos et al., 1992), murine (Miller, 1992; Miller et al., 1985; Sorge et al., 1984; Mann and Baltimore, 1985; Miller et al., 1988), and human origin (Shimada et al., 1991; Helseth et al., 1990; Page et al., 1990; Buchschacher and Panganiban, 1992). Most human gene therapy protocols have been based on disabled murine retroviruses.

Nonviral gene transfer methods known in the art include chemical techniques such as calcium phosphate coprecipitation (Graham and van der Eb, 1973; Pellicer et al., 1980); mechanical techniques, for example microinjection (Anderson et al., 1980; Gordon et al., 1980; Brinster et al., 1981; Constantini and Lacy, 1981); membrane fusion-mediated transfer via liposomes (Felgner et al., 1987; Wang and Huang, 1989; Kaneda et al., 1989; Stewart et al., 1992; Nabel et al., 1990; Lim et al., 1992); and direct DNA uptake and receptor-mediated DNA transfer (Wolff et al., 1990; Wu et al., 1991; Zenke et al., 1990; Wu et al., 1989b; Wolff et al., 1991; Wagner et al., 1990; Wagner et al., 1991; Cotten et al., 1990; Curiel et al., 1991a; Curiel et al., 1991b). Viral-mediated gene transfer can be combined with direct in vivo gene transfer using liposome delivery, allowing one to direct the viral vectors to the tumor cells and not into the surrounding nondividing cells. Alternatively, the retroviral vector producer cell line can be injected into tumors (Culver et al., 1992). Injection of producer cells would then provide a continuous source of vector particles. This technique has been approved for use in humans with inoperable brain tumors.

In an approach which combines biological and physical gene transfer methods, plasmid DNA of any size is combined with a polylysine-conjugated antibody specific to the adenovirus hexon protein, and the resulting complex is bound to an adenovirus vector. The trimolecular complex is then used to infect cells. The adenovirus vector permits efficient binding, internalization, and degradation of the endosome before the coupled DNA is damaged.

Liposome/DNA complexes have been shown to be capable of mediating direct in vivo gene transfer. While in

standard liposome preparations the gene transfer process is nonspecific, localized in vivo uptake and expression have been reported in tumor deposits, for example, following direct in situ administration (Nabel, 1992).

Gene transfer techniques which target DNA directly to breast and ovarian tissues, e.g., epithelial cells of the breast or ovaries, is preferred. Receptor-mediated gene transfer, for example, is accomplished by the conjugation of DNA (usually in the form of covalently closed supercoiled plasmid) to a protein ligand via polylysine. Ligands are chosen on the basis of the presence of the corresponding ligand receptors on the cell surface of the target cell/tissue type. One appropriate receptor/ligand pair may include the estrogen receptor and its ligand, estrogen (and estrogen analogues). These ligand-DNA conjugates can be injected directly into the blood if desired and are directed to the target tissue where receptor binding and internalization of the DNA-protein complex occurs. To overcome the problem of intracellular destruction of DNA, coinfection with adenovirus can be included to disrupt endosome function.

The therapy involves two steps which can be performed singly or jointly. In the first step, prepubescent females who carry a BRCA1 susceptibility allele are treated with a gene delivery vehicle such that some or all of their mammary ductal epithelial precursor cells receive at least one additional copy of a functional normal BRCA1 allele. In this step, the treated individuals have reduced risk of breast cancer to the extent that the effect of the susceptible allele has been countered by the presence of the normal allele. In the second step of a preventive therapy, predisposed young females, in particular women who have received the proposed gene therapeutic treatment, undergo hormonal therapy to mimic the effects on the breast of a full term pregnancy. Methods of Use: Peptide Therapy

Peptides which have BRCA1 activity can be supplied to cells which carry mutant or missing BRCA1 alleles. The sequence of the BRCA1 protein is disclosed (SEQ ID NO:2). Protein can be produced by expression of the cDNA sequence in bacteria, for example, using known expression vectors. Alternatively, BRCA1 polypeptide can be extracted from BRCA1-producing mammalian cells. In addition, the techniques of synthetic chemistry can be employed to synthesize BRCA1 protein. Any of such techniques can provide the preparation of the present invention which comprises the BRCA1 protein. The preparation is substantially free of other human proteins. This is most readily accomplished by synthesis in a microorganism or in vitro.

Active BRCA1 molecules can be introduced into cells by microinjection or by use of liposomes, for example. Alternatively, some active molecules may be taken up by cells, actively or by diffusion. Extracellular application of the BRCA1 gene product may be sufficient to affect tumor growth. Supply of molecules with BRCA1 activity should lead to partial reversal of the neoplastic state. Other molecules with BRCA1 activity (for example, peptides, drugs or organic compounds) may also be used to effect such a reversal. Modified polypeptides having substantially similar function are also used for peptide therapy.

Methods of Use: Transformed Hosts

Similarly, cells and animals which carry a mutant BRCA1 allele can be used as model systems to study and test for substances which have potential as therapeutic agents. The cells are typically cultured epithelial cells. These may be isolated from individuals with BRCA1 mutations, either somatic or germline. Alternatively, the cell line can be engineered to carry the mutation in the BRCA1 allele, as described above. After a test substance is applied to the cells,

the neoplastically transformed phenotype of the cell is determined. Any trait of neoplastically transformed cells can be assessed, including anchorage-independent growth, tumorigenicity in nude mice, invasiveness of cells, and growth factor dependence. Assays for each of these traits are known in the art.

Animals for testing therapeutic agents can be selected after mutagenesis of whole animals or after treatment of germline cells or zygotes. Such treatments include insertion of mutant BRCA1 alleles, usually from a second animal species, as well as insertion of disrupted homologous genes. Alternatively, the endogenous BRCA1 gene(s) of the animals may be disrupted by insertion or deletion mutation or other genetic alterations using conventional techniques (Capecci, 1989; Valancius and Smithies, 1991; Hasty et al., 1991; Shinkai et al., 1992; Mombaerts et al., 1992; Philpott et al., 1992; Snouwaert et al., 1992; Donehower et al., 1992). After test substances have been administered to the animals, the growth of tumors must be assessed. If the test substance prevents or suppresses the growth of tumors, then the test substance is a candidate therapeutic agent for the treatment of the cancers identified herein. These animal models provide an extremely important testing vehicle for potential therapeutic products.

The present invention is described by reference to the following Examples, which are offered by way of illustration and are not intended to limit the invention in any manner. Standard techniques well known in the art or the techniques specifically described below were utilized.

EXAMPLE 1

Ascertain and Study Kindreds Likely to Have a 1
7q-Linked Breast Cancer Susceptibility Locus

Extensive cancer prone kindreds were ascertained by our University of Utah collaborators from a defined population providing a large set of extended kindreds with multiple cases of breast cancer and many relatives available to study. The large number of meioses present in these large kindreds provided the power to detect whether the BRCA1 locus was segregating, and increased the opportunity for informative recombinants to occur within the small region being investigated. This vastly improved the chances of establishing linkage to the BRCA1 region, and greatly facilitated the

reduction of the BRCA1 region to a manageable size, which permits identification of the BRCA1 locus itself.

Each kindred was extended through all available connecting relatives by our collaborators, and to all informative first degree relatives of each proband or cancer case. For these kindreds, additional breast cancer cases and individuals with cancer at other sites of interest (e.g. ovarian) who also appeared in the kindreds were identified through the tumor registry linked files. All breast cancers reported in the kindred which were not confirmed in the Utah Cancer Registry were researched. Medical records or death certificates were obtained for confirmation of all cancers. Each key connecting individual and all informative individuals were invited by our collaborators to participate by providing a blood sample from which DNA was extracted. They also sampled spouses and relatives of deceased cases so that the genotype of the deceased cases could be inferred from the genotypes of their relatives.

Ten kindreds which had three or more cancer cases with inferable genotypes were selected for linkage studies to 17q markers from a set of 29 kindreds originally ascertained for a study of proliferative breast disease and breast cancer (Skolnick et al., 1990). The criterion for selection of these kindreds was the presence of two sisters or a mother and her daughter with breast cancer. Additionally, two kindreds which have been studied since 1980 as part of their breast cancer linkage studies (K1001, K9018), six kindreds ascertained from for the presence of clusters of breast and/or ovarian cancer (K2019, K2073, K2079, K2080, K2039, K2082) and a self-referred kindred with early onset breast cancer (K2035) were included. These kindreds were investigated and expanded in our collaborators clinic in the manner described above. Table 1 displays the characteristics of these 19 kindreds which are the subject of subsequent exome each kin Table 1, for each kindred the total number of individuals in our database, the number of typed individuals, and the minimum, median, and maximum age at diagnosis of breast/ovarian cancer are reported. Kindreds are sorted in ascending order of median age at diagnosis of breast cancer. Four women diagnosed with both ovarian and breast cancer are counted in both categories.

TABLE 1

KINDRED	Description of the 19 Kindreds									
	No. of		Breast				Ovarian			
			Individuals		Age at Dx		Age at Dx		Age at Dx	
	Total	Sample	# Aff.	Min.	Med.	Max.	# Aff.	Min.	Med.	Max.
1910	15	10	4	27	34	49	—	—	—	—
1001	133	98	13	28	37	64	—	—	—	—
2035	42	25	8	28	37	45	1	—	60	—
2027	21	11	4	34	38	41	—	—	—	—
9018	54	17	9	30	40	72	2	46	48	50
1925	50	27	4	39	42	53	—	—	—	—
1927	49	29	5	32	42	51	—	—	—	—
1911	28	21	7	28	42	76	—	—	—	—
1929	16	11	4	34	43	73	—	—	—	—
1901	35	19	10	31	44	76	—	—	—	—
2082	180	105	20	27	47	67	10	45	52	66
2019	42	19	10	42	53	79	—	—	—	—

5,753,441

37

38

TABLE 1-continued

Description of the 19 Kindreds										
No. of		Breast					Ovarian			
		Age at Dx					Age at Dx			
KINDRED	Total	Sample	# Aff.	Min.	Med.	Max.	# Aff.	Min.	Med.	Max.
1900	70	23	8	45	55	70	1	—	78	—
2080	264	74	22+	27	55	92	4	45	53	71
2073	57	29	9	35	57	80	—	—	—	—
1917	16	6	4	43	58	61	—	—	—	—
1920	22	14	3	62	63	68	—	—	—	—
2079	136	18	14	38	66	84	4	52	59	65
2039	87	40	14	44	68	88	4	41	51	75

+Includes one case of male breast cancer.

EXAMPLE 2

Selection of Kindreds Which are Linked to Chromosome 17q and Localization of BRCA1 to the Interval Mfd15–Mfd188

For each sample collected in these 19 kindreds, DNA was extracted from blood (or in two cases from paraffin-embedded tissue blocks) using standard laboratory protocols. Genotyping in this study was restricted to short tandem repeat (STR) markers since, in general, they have high heterozygosity and PCR methods offer rapid turnaround while using very small amounts of DNA. To aid in this effort, four such STR markers on chromosome 17 were developed by screening a chromosome specific cosmid library for CA positive clones. Three of these markers localized to the long arm: (46E6, Easton et al., 1993); (42D6, Easton et al., 1993); 26C2 (D17S514, Oliphant et al., 1991), while the other, 12G6 (D17S513, Oliphant et al., 1991), localized to the short arm near the p53 tumor suppressor locus. Two of these, 42D6 and 46E6, were submitted to the Breast Cancer Linkage Consortium for typing of breast cancer families by investigators worldwide. Oligonucleotide sequences for markers not developed in our laboratory were obtained from published reports, or as part of the Breast Cancer Linkage Consortium, or from other investigators. All genotyping films were scored blindly with a standard lane marker used to maintain consistent coding of alleles. Key

samples in the four kindreds presented here underwent duplicate typing for all relevant markers. All 19 kindreds have been typed for two polymorphic CA repeat markers: 42D6 (D17S588), a CA repeat isolated in our laboratory, and Mfd15 (D17S250), a CA repeat provided by J. Weber (Weber et al., 1990). Several sources of probes were used to create genetic markers on chromosome 17, specifically chromosome 17 cosmid and lambda phage libraries created from sorted chromosomes by the Los Alamos National Laboratories (van Dilla et al., 1986).

LOD scores for each kindred with these two markers (42D6, Mfd15) and a third marker, Mfd188 (D17S579, Hall et al., 1992), located roughly midway between these two markers, were calculated for two values of the recombination fraction, 0.001 and 0.1. (For calculation of LOD scores, see Oh, 1985). Likelihoods were computed under the model derived by Claus et al., 1991, which assumes an estimated gene frequency of 0.003, a lifetime risk in gene carriers of about 0.80, and population based age-specific risks for breast cancer in non-gene carriers. Allele frequencies for the three markers used for the LOD score calculations were calculated from our own laboratory typings of unrelated individuals in the CEPH panel (White and Lalouel, 1988). Table 2 shows the results of the pairwise linkage analysis of each kindred with the three markers 42D6, Mfd188, and Mfd15.

TABLE 2

Pairwise Linkage Analysis of Kindreds						
KINDRED	Mfd15 (D17S250) Recombination		Mfd188 (D17S579) Recombination		42D6 (D17S588) Recombination	
	0.001	0.1	0.001	0.1	0.001	0.1
1910	0.06	0.30	0.06	0.30	0.06	0.30
1001	-0.30	-0.09	NT	NT	-0.52	-0.19
2035	2.34	1.85	0.94	0.90	2.34	1.82
2027	-1.22	-0.33	-1.20	-0.42	-1.16	-0.33
9018	-0.54	-0.22	-0.17	-0.10	0.11	0.07
1925	1.08	0.79	0.55	0.38	-0.11	-0.07
1927	-0.41	0.01	-0.35	0.07	-0.44	-0.02
1911	-0.27	-0.13	-0.43	-0.23	0.49	0.38
1929	-0.49	-0.25	NT	NT	-0.49	-0.25
1901	1.50	1.17	0.78	0.57	0.65	0.37
2082	4.25	3.36	6.07	5.11	2.00	3.56
2019	-0.10	-0.01	-0.11	-0.05	-0.18	-0.10

TABLE 2-continued

KINDRED	Pairwise Linkage Analysis of Kindreds					
	Mfd15 (D17S250)		Mfd188 (D17S579)		42D6 (D17S588)	
	Recombination		Recombination		Recombination	
	0.001	0.1	0.001	0.1	0.001	0.1
1900	-0.14	-0.11	NT	NT	-0.12	-0.05
2080	-0.16	-0.04	0.76	0.74	-1.25	-0.58
2073	-0.41	-0.29	0.63	0.49	-0.23	-0.13
1917	-0.02	-0.02	NT	NT	-0.01	0.00
1920	-0.03	-0.02	NT	NT	0.00	0.00
2079	0.02	0.01	-0.01	-0.01	0.01	0.01
2039	-1.67	-0.83	0.12	0.59	-1.15	0.02

NT - Kindred not typed for Mfd188.

Using a criterion for linkage to 17q of a LOD score>1.0 for at least one locus under the CASH model (Claus et al., 1991), four of the 19 kindreds appeared to be linked to 17q (K1901, K1925, K2035, K2082). A number of additional kindreds showed some evidence of linkage but at this time could not be definitively assigned to the linked category. These included kindreds K1911, K2073, K2039, and K2080. Three of the 17q-linked kindreds had informative recombinants in this region and these are detailed below.

Kindred 2082 is the largest 17q-linked breast cancer family reported to date by any group. The kindred contains 20 cases of breast cancer, and ten cases of ovarian cancer. Two cases have both ovarian and breast cancer. The evidence of linkage to 17q for this family is overwhelming; the LOD score with the linked haplotype is over 6.0, despite the existence of three cases of breast cancer which appear to be sporadic, i.e., these cases share no part of the linked haplotype between Mfd15 and 42D6. These three sporadic cases were diagnosed with breast cancer at ages 46, 47, and 54. In smaller kindreds, sporadic cancers of this type greatly confound the analysis of linkage and the correct identification of key recombinants. The key recombinant in the 2082 kindred is a woman who developed ovarian cancer at age 45 whose mother and aunt had ovarian cancer at ages 58 and 66, respectively. She inherited the linked portion of the haplotype for both Mfd188 and 42D6 while inheriting unlinked alleles at Mfd15; this recombinant event placed BRCA1 distal to Mfd15.

K1901 is typical of early-onset breast cancer kindreds. The kindred contains 10 cases of breast cancer with a median age at diagnosis of 43.5 years of age; four cases were diagnosed under age 40. The LOD score for this kindred with the marker 42D6 is 1.5, resulting in a posterior probability of 17q-linkage of 0.96. Examination of haplotypes in this kindred identified a recombinant haplotype in an obligate male carrier and his affected daughter who was diagnosed with breast cancer at age 45. Their linked allele for marker Mfd15 differs from that found in all other cases in the kindred (except one case which could not be completely inferred from her children). The two haplotypes are identical for Mfd188 and 42D6. Accordingly, data from Kindred 1901 would also place the BRCA1 locus distal to Mfd15.

Kindred 2035 is similar to K1901 in disease phenotype. The median age of diagnosis for the eight cases of breast cancer in this kindred is 37. One case also had ovarian cancer at age 60. The breast cancer cases in this family descend from two sisters who were both unaffected with breast cancer until their death in the eighth decade. Each branch contains four cases of breast cancer with at least one case in each branch having markedly early onset. This

kindred has a LOD score of 2.34 with Mfd15. The haplotypes segregating with breast cancer in the two branches share an identical allele at Mfd15 but differ for the distal loci Mfd188 and NM23 (a marker typed as part of the consortium which is located just distal to 42D6 (Hall et al., 1992)). Although the two haplotypes are concordant for marker 42D6, it is likely that the alleles are shared identical by state (the same allele but derived from different ancestors), rather than identical by descent (derived from a common ancestor) since the shared allele is the second most common allele observed at this locus. By contrast the linked allele shared at Mfd15 has a frequency of 0.04. This is a key recombinant in our dataset as it is the sole recombinant in which BRCA1 segregated with the proximal portion of the haplotype, thus setting the distal boundary to the BRCA1 region. For this event not to be a key recombinant requires that a second mutant BRCA1 gene be present in a spouse marrying into the kindred who also shares the rare Mfd15 allele segregating with breast cancer in both branches of the kindred. This event has a probability of less than one in a thousand. The evidence from this kindred therefore placed the BRCA1 locus proximal to Mfd188.

EXAMPLE 3

Creation of a Fine Structure Map and Refinement of the BRCA1 Region to Mfd191-Mfd188 using Additional STR Polymorphisms

In order to improve the characterization of our recombinants and define closer flanking markers, a dense map of this relatively small region on chromosome 17q was required. The chromosome 17 workshop has produced a consensus map of this region (FIG. 1) based on a combination of genetic and physical mapping studies (Fain, 1992). This map contains both highly polymorphic STR polymorphisms, and a number of nonpolymorphic expressed genes. Because this map did not give details on the evidence for this order nor give any measure of local support for inversions in the order of adjacent loci, we viewed it as a rough guide for obtaining resources to be used for the development of new markers and construction of our own detailed genetic and physical map of a small region containing BRCA1. Our approach was to analyze existing STR markers provided by other investigators and any newly developed markers from our laboratory with respect to both a panel of meiotic (genetic) breakpoints identified using DNA from the CEPH reference families and a panel of somatic cell hybrids (physical breakpoints) constructed for this region. These markers included 26C2 developed in our laboratory which maps proximal to Mfd15, Mfd191 (provided by James Weber),

5,753,441

41

THRA1 (Futreal et al., 1992a), and three polymorphisms kindly provided to us by Dr. Donald Black, NM23 (Hall et al. 1992), SCG40 (D17S181), and 6C1 (D17S293).

Genetic localization of markers.

In order to localize new markers genetically within the region of interest, we have identified a number of key meiotic breakpoints within the region, both in the CEPH reference panel and in our large breast cancer kindred (K2082). Given the small genetic distance in this region, they are likely to be only a relatively small set of recombinants which can be used for this purpose, and they are likely to group markers into sets. The orders of the markers within each set can only be determined by physical mapping. However the number of genotypings necessary to position a new marker is minimized. These breakpoints are illustrated in Tables 3 and 4. Using this approach we were able to genetically order the markers THRA1, 6C1, SCG40, and Mfd191. As can be seen from Tables 3 and 4, THRA1 and MFD191 both map inside the Mfd15–Mfd188 region we had previously identified as containing the BRCA1 locus. In Tables 3 and 4, M/P indicates a maternal or paternal recombinant. A “1” indicates inherited allele is of grandpaternal origin, while a “0” indicates grandmaternal origin, and “-” indicates that the locus was untyped or uninformative.

TABLE 3

CEPH Recombinants									
Family	ID	M/P	Mfd15	THRA1	Mfd191	Mfd188	SCG40	6C1	42D6
13292	4	M	1	1	1	0	0	0	0
13294	4	M	1	1	1	0	0	0	0
13294	6	M	0	0	1	1	—	—	—
1334	3	M	1	1	1	1	1	0	0
1333	4	M	1	1	1	0	—	—	0
1333	6	M	0	0	1	1	—	—	1
1333	8	P	1	0	0	0	—	—	0
1377	8	M	0	—	0	0	0	0	1

TABLE 4

Kindred 2082 Recombinants							
Family	ID	M/P	Mfd15	Mfd191	Mfd188	SCG40	6C1
75		M	0	1	1	1	—
63		M	0	0	1	1	—
125		M	1	1	1	0	—
40		M	1	1	0	0	—

Analysis of markers Mfd15, Mfd188, Mfd191, and THRA1 in our recombinant families.

Mfd15, Mfd188, Mfd191 and THRA1 were typed in our recombinant families and examined for additional information to localize the BRCA1 locus. In kindred 1901, the Mfd15 recombinant was recombinant for THRA1 but uninformative for Mfd191, thus placing BRCA1 distal to THRA1. In K2082, the recombinant with Mfd15 also was recombinant with Mfd191, thus placing the BRCA1 locus distal to Mfd191 (Goldgar et al., 1994). Examination of THRA1 and Mfd191 in kindred K2035 yielded no further localization information as the two branches were concordant for both markers. However, SCG40 and 6C1 both displayed the same pattern as Mfd188, thus increasing our confidence in the localization information provided by the Mfd188 recombinant in this family. The BRCA1 locus, or at

42

least a portion of it, therefore lies within an interval bounded by Mfd191 on the proximal side and Mfd188 on the distal side.

EXAMPLE 4

Development of Genetic and Physical Resources in the Region of Interest

To increase the number of highly polymorphic loci in the Mfd191–Mfd188 region, we developed a number of STR markers in our laboratory from cosmids and YACs which physically map to the region. These markers allowed us to further refine the region.

STSs were identified from genes known to be in the desired region to identify YACs which contained these loci, which were then used to identify subclones in cosmids, PIs or BACs. These subclones were then screened for the presence of a CA tandem repeat using a (CA)_n oligonucleotide (Pharmacia). Clones with a strong signal were selected preferentially, since they were more likely to represent CA-repeats which have a large number of repeats and/or are of near-perfect fidelity to the (CA)_n pattern. Both of these characteristics are known to increase the probability of polymorphism (Weber, 1990). These clones were sequenced

directly from the vector to locate the repeat. We obtained a unique sequence on one side of the CA-repeat by using one of a set of possible primers complementary to the end of a CA-repeat, such as (GT)₁₀T. Based on this unique sequence, a primer was made to sequence back across the repeat in the other direction, yielding a unique sequence for design of a second primer flanking the CA-repeat. STRs were then screened for polymorphism on a small group of unrelated individuals and tested against the hybrid panel to confirm their physical localization. New markers which satisfied these criteria were then typed in a set of 40 unrelated individuals from the Utah and CEPH families to obtain allele frequencies appropriate for the study population. Many of the other markers reported in this study were tested in a smaller group of CEPH unrelated individuals to obtain similarly appropriate allele frequencies.

Using the procedure described above, a total of eight polymorphic STRs was found from these YACS. Of the loci identified in this manner, four were both polymorphic and localized to the BRCA1 region. Four markers did not localize to chromosome 17, reflecting the chimeric nature of the YACs used. The four markers which were in the region were denoted AA1, ED2, 4–7, and YM29. AA1 and ED2 were developed from YACs positive for the RNU2 gene, 4–7 from an EPB3 YAC and YM29 from a cosmid which localized to the region by the hybrid panel. A description of the number of alleles, heterozygosity and source of these

5,753,441

43

four and all other STR polymorphisms analyzed in the breast cancer kindreds is given below in Table 5.

TABLE 5

Polymorphic Short Tandem Repeat Markers Used for Fine Structure Mapping of the BRCA1 Locus										
Clone	Gene	Na**	Heterozygosity	Allele* Frequency (%)						
				1	2	3	4	5	6	
Mfd15	D17S250	10	0.82	26	22	15	7	7	23	
THRA1	THRA1	5								
Mfd191	D17S776	7	0.55	48	20	11	7	7	7	
ED2	D17S1327	12	0.55	62	9	8	5	5	11	
AA1	D17S1326	7	0.83	28	28	25	8	6	5	
CA375	D17S184	10	0.75	26	15	11	9	9	20	
4-7	D17S1183	9	0.50	63	15	8	6	4	4	
YM29	—	9	0.62	42	24	12	7	7	8	
Mfd188	D17S579	12	0.92	33	18	8	8	8	25	
SCG40	D17S181	14	0.90	20	18	18	10	8	35	
42D6	D17S588	11	0.86	21	17	11	10	9	32	
6C1	D17S293	7	0.75	30	30	11	11	9	9	
Z109	D17S750	9	0.70	33	27	7	7	7	19	
tdj1475	D17S1321	13	0.84	21	16	11	11	8	33	
CF4	D17S1320	6	0.63	50	27	9	7	4	3	
tdj1239	D17S1328	10	0.80	86	10	9	7	4	14	
U5	D17S1325	13	0.83	19	16	12	10	9	34	

*Allele codes 1-5 are listed in decreasing frequency; allele numbers do not correspond to fragment sizes. Allele 6 frequency is the joint frequency of all other alleles for each locus.
**Number of alleles seen in the genetically independent DNA samples used for calculating allele frequencies.

The four STR polymorphisms which mapped physically to the region (4-7, ED2, AA1, YM29) were analyzed in the meiotic, breakpoint panel shown initially in Tables 3 and 4. Tables 6 and 7 contain the relevant CEPH data and Kindred 2082 data for localization of these four markers. In the tables, M/P indicates a maternal or paternal recombinant. A "1" indicates inherited allele is of grandpaternal origin, while a "0" indicates grandmaternal origin, and "-" indicates that the locus was untyped or uninformative.

TABLE 6

Key Recombinants Used for Genetic Ordering of New STR Loci Developed in Our Laboratory Within the BRCA1 Region of 17q													
CEPH Family	ID	M/P	Mfd15	THRA1	Mfd191	ED2	AA1	Z109	4-7	YM29	Mfd188	SCG40	42D6
13292	4	M	1	1	1	1	1	0	0	0	0	0	0
13294	4	M	1	0	0	—	0	—	—	—	0	—	—
13294	6	M	0	0	1	—	1	—	—	—	1	—	—
1333	4	M	1	1	1	—	0	—	—	0	0	—	0
1333	6	M	0	0	1	—	1	—	—	1	1	—	1
1333	3	M	0	0	1	—	—	—	1	1	1	—	1

TABLE 7

Kindred 2082 Recombinants										
ID	M/P	Mfd15	Mfd191	ED2	AA1	4-7	YM29	Mfd188	SCG40	42D6
63	M	0	0	1	—	1	1	1	1	1
125	M	1	1	1	—	1	1	1	0	0
40	M	1	1	0	—	0	—	0	0	0
22	P	0	0	1	1	1	1	1	1	1

44

From CEPH 1333-04, we see that AA1 and YM29 must lie distal to Mfd191. From 13292, it can be inferred that both AA1 and ED2 are proximal to 4-7, YM29, and Mfd188. The recombinants found in K2082 provide some additional ordering information. Three independent observations (individual numbers 22, 40, & 63) place AA1, ED2, 4-7, and YM29, and Mfd188 distal to Mfd191, while ID 125 places 4-7, YM29, and Mfd188 proximal to SCG40. No genetic information on the relative ordering within the two clusters of markers AA1/ED2 and 4-7/YM29/Mfd188 was obtained from the genetic recombinant analysis. Although ordering loci with respect to hybrids which are known to contain "holes" in which small pieces of interstitial human DNA may be missing is problematic, the hybrid patterns indicate that 4-7 lies above both YM29 and Mfd188.

EXAMPLE 5

Genetic Analyses of Breast Cancer Kindreds with Markers AA1, 4-7, ED2, and YM29

In addition to the three kindreds containing key recombinants which have been discussed previously, kindred K2039 was shown through analysis of the newly developed STR markers to be linked to the region and to contain a useful recombinant.

Table 8 defines the haplotypes (shown in coded form) of the kindreds in terms of specific marker alleles at each locus and their respective frequencies. In Table 8, alleles are listed in descending order of frequency; frequencies of alleles 1-5 for each locus are given in Table 5. Haplotypes coded H are BRCA1 associated haplotypes, P designates a partial H haplotype, and an R indicates an observable recombinant haplotype. As evident in Table 8, not all kindreds were typed for all markers; moreover, not all individuals within a kindred were typed for an identical set of markers, especially in K2082. With one exception, only haplotypes inherited from affected or at-risk kindred members are shown; haplotypes from spouses marrying into the kindred are not

5,753,441

45

described. Thus in a given sibship, the appearance of haplotypes X and Y indicates that both haplotypes from the affected/at-risk individual were seen and neither was a breast cancer associated haplotype.

46

affected daughter. In this family the odds in favor of affected individuals carrying BRCA1 susceptibility alleles are extremely high; the only possible interpretations of the data are that BRCA1 is distal to Mfd191 or alternatively that the

TABLE 8

Breast Cancer Linked Haplotypes Found in the Three Kindreds														
Kin.	HAP	Mfd 15	THRA1	Mfd 191	tdj 1475	ED2	AA1	Z109	CA375	4-7	YM29	Mfd 188	SCG40	6C1 42D6
1901	H1	1	5	5	3	1	4	NI	NI	1	1	3	NI	NI 1
	R2	9	2	5	6	1	4	NI	NI	1	1	3	NI	NI 1
	H1	3	NI	4	6	6	1	NI	NI	2	1	4	2	NI 1
	P1	3	NI	4	NI	NI	NI	NI	NI	NI	NI	4	2	NI 1
2082	P2	3	NI	NI	NI	NI	NI	NI	NI	NI	NI	4	NI	NI NI
	R1	6	NI	1	5	6	1	NI	NI	2	1	4	2	NI 1
	R2	6	NI	4	6	6	1	NI	NI	2	1	4	2	NI 1
	R3	3	NI	4	NI	6	1	NI	NI	2	1	4	1	NI 7
	R4	7	NI	1	NI	1	5	NI	NI	4	6	1	2	NI 1
	R5	3	NI	4	NI	NI	NI	NI	NI	NI	2	1	NI	NI NI
	R6	3	NI	4	3	1	2	NI	NI	1	2	2	6	NI 6
	R7	3	NI	4	3	7	1	NI	NI	1	1	3	7	NI 4
	H1	8	2	1	NI	5	1	1	4	3	1	6	8	2 4
	H2	8	2	1	NI	5	1	1	2	1	1	2	3	1 4
2035	R2	8	2	1	NI	5	1	1	2	1	1	2	3	6 1

In kindred K1901, the new markers showed no observable recombination with breast cancer susceptibility, indicating that the recombination event in this kindred most likely took place between THRA1 and ED2. Thus, no new BRCA1 localization information was obtained based upon studying the four new markers in this kindred. In kindred 2082 the key recombinant individual has inherited the linked alleles for ED2, 4-7, AA1, and YM29, and was recombinant for tdj1474 indicating that the recombination event occurred in this individual between tdj1474 and ED2/AA1.

There are three haplotypes of interest in kindred K2035, H1, H2, and R2 shown in Table 8. H1 is present in the four cases and one obligate male carrier descendant from individual 17 while H2 is present or inferred in two cases and two obligate male carriers in descendants of individual 10. R2 is identical to H2 for loci between and including Mfd15 and SCG40, but has recombined between SCG40 and 42D6. Since we have established that BRCA1 is proximal to 42D6, this H2/R2 difference adds no further localization information. H1 and R2 share an identical allele at Mfd15, THRA1, AA1, and ED2 but differ for loci presumed distal to ED2, i.e., 4-7, Mfd188, SCG40, and 6C1. Although the two haplotypes are concordant for the 5th allele for marker YM29, a marker which maps physically between 4-7 and Mfd188, it is likely that the alleles are shared identical by state rather than identical by descent since this allele is the most common allele at this locus with a frequency estimated in CEPH parents of 0.42. By contrast, the linked alleles shared at the Mfd15 and ED2 loci have frequencies of 0.04 and 0.09, respectively. They also share more common alleles at Mfd191 (frequency=0.52), THRA1, and AA1 (frequency=0.28). This is the key recombinant in the set as it is the sole recombinant in which breast cancer segregated with the proximal portion of the haplotype, thus setting the distal boundary. The evidence from this kindred therefore places the BRCA1 locus proximal to 4-7.

The recombination event in kindred 2082 which places BRCA1 distal to tdj1474 is the only one of the four events described which can be directly inferred; that is, the affected mother's genotype can be inferred from her spouse and offspring, and the recombinant haplotype can be seen in her

purported recombinant is a sporadic case of ovarian cancer at age 44. Rather than a directly observable or inferred recombinant, interpretation of kindred 2035 depends on the observation of distinct 17q-haplotypes segregating in different and sometimes distantly related branches of the kindred. The observation that portions of these haplotypes have alleles in common for some markers while they differ at other markers places the BRCA1 locus in the shared region. The confidence in this placement depends on several factors: the relationship between the individuals carrying the respective haplotypes, the frequency of the shared allele, the certainty with which the haplotypes can be shown to segregate with the BRCA1 locus, and the density of the markers in the region which define the haplotype. In the case of kindred 2035, the two branches are closely related, and each branch has a number of early onset cases which carry the respective haplotype. While two of the shared alleles are common, (Mfd191, THRA1), the estimated frequencies of the shared alleles at Mfd15, AA1, and ED2 are 0.04, 0.28, and 0.09, respectively. It is therefore highly likely that these alleles are identical by descent (derived from a common ancestor) rather than identical by state (the same allele but derived from the general population).

EXAMPLE 6

Refined Physical Mapping Studies Place the BRCA1 Gene in a Region Flanked by tdj1474 and U5R

Since its initial localization to chromosome 17q in 1990 (Hall et al., 1990) a great deal of effort has gone into localizing the BRCA1 gene to a region small enough to allow implementation of effective positional cloning strategies to isolate the gene. The BRCA1 locus was first localized to the interval Mfd15 (D17S250)-42D6 (D17S588) by multipoint linkage analysis (Easton et al., 1993) in the collaborative Breast Cancer Linkage Consortium dataset consisting of 214 families collected worldwide. Subsequent refinements of the localization have been based upon individual recombinant events in specific families. The region THRA1-D17S183 was defined by Bowcock et al., 1993; and the region THRA1-D17S78 was defined by Simard et al., 1993.

5,753,441

47

We further showed that the BRCA1 locus must lie distal to the marker Mfd191 (D17S776) (Goldgar et al., 1994). This marker is known to lie distal to THRA1 and RARA. The smallest published region for the BRCA1 locus is thus between D17S776 and D17S78. This region still contains approximately 1.5 million bases of DNA, making the isolation and testing of all genes in the region a very difficult task. We have therefore undertaken the tasks of constructing a physical map of the region, isolating a set of polymorphic STR markers located in the region, and analyzing these new markers in a set of informative families to refine the location of the BRCA1 gene to a manageable interval.

Four families provide important genetic evidence for localization of BRCA1 to a sufficiently small region for the application of positional cloning strategies. Two families (K2082, K1901) provide data relating to the proximal boundary for BRCA1 and the other two (K2035, K1813) fix the distal boundary. These families are discussed in detail below. A total of 15 Short Tandem Repeat markers assayable by PCR were used to refine this localization in the families studied. These markers include DS17S7654, DS17S975, tdj1474, and tdj1239. Primer sequences for these markers are provided in SEQ ID NO:3 and SEQ ID NO:4 for DS17S754; in SEQ ID NO:5 and SEQ ID NO:6 for DS17S975; in SEQ ID NO:7 and SEQ ID NO:8 for tdj1474; and, in SEQ ID NO:9 and SEQ ID NO:10 for tdj1239.

Kindred 2082

Kindred 2082 is the largest BRCA1-linked breast/ovarian cancer family studied to date. It has a LOD score of 8.6, providing unequivocal evidence for 17q linkage. This family has been previously described and shown to contain a critical recombinant placing BRCA1 distal to MFD191 (D17S776). This recombinant occurred in a woman diagnosed with ovarian cancer at age 45 whose mother had ovarian cancer at age 63. The affected mother was deceased; however, from her children, she could be inferred to have the linked haplotype present in the 30 other linked cases in the family in the region between Mfd15 and Mfd188. Her affected daughter received the linked allele at the loci ED2, 4-7, and Mfd188, but received the allele on the non-BRCA1 chromosome at Mfd15 and Mfd191. In order to further localize this recombination breakpoint, we tested DNA from the key members of this family for the following markers derived from physical mapping resources: tdj1474, tdj1239, CF4, D17S855. For the markers tdj1474 and CF4, the affected daughter did not receive the linked allele. For the STR locus tdj1239, however, the mother could be inferred to be informative and her daughter did receive the BRCA1-associated allele. D17S855 was not informative in this family. Based on this analysis, the order is 17q centromere—Mfd191—17HSD—CF4—tdj1474—tdj1239—D17S855—ED2—4-7—Mfd188—17q telomere. The recombinant described above therefore places BRCA1 distal to tdj1474, and the breakpoint is localized to the interval between tdj1474 and tdj1239. The only alternative explanation for the data in this family other than that of BRCA1 being located distal to tdj1474, is that the ovarian cancer present in the recombinant individual is caused by reasons independent of the BRCA1 gene. Given that ovarian cancer diagnosed before age 50 is rare, this alternate explanation is exceedingly unlikely.

Kindred 1901

Kindred 1901 is an early-onset breast cancer family with 7 cases of breast cancer diagnosed before 50, 4 of which were diagnosed before age 40. In addition, there were three cases of breast cancer diagnosed between the ages of 50 and 70. One case of breast cancer also had ovarian cancer at age

48

61. This family currently has a LOD score of 1.5 with D17S855. Given this linkage evidence and the presence of at least one ovarian cancer case, this family has a posterior probability of being due to BRCA1 of over 0.99. In this family, the recombination comes from the fact that an individual who is the brother of the ovarian cancer case from which the majority of the other cases descend, only shares a portion of the haplotype which is cosegregating with the other cases in the family. However, he passed this partial haplotype to his daughter who developed breast cancer at age 44. If this case is due to the BRCA1 gene, then only the part of the haplotype shared between this brother and his sister can contain the BRCA1 gene. The difficulty in interpretation of this kind of information is that while one can be sure of the markers which are not shared and therefore recombinant, markers which are concordant can either be shared because they are non-recombinant, or because their parent was homozygous. Without the parental genotypic data it is impossible to discriminate between these alternatives. Inspection of the haplotype in K1901, shows that he does not share the linked allele at Mfd15 (D17S250), THRA1, CF4 (D17S1320), and tdj1474 (17DS1321). He does share the linked allele at Mfd191 (D17S776), ED2 (D17S1327), tdj1239 (D17S1328), and Mfd188 (D17S579). Although the allele shared at Mfd191 is relatively rare (0.07), we would presume that the parent was homozygous since they are recombinant with markers located nearby on either side, and a double recombination event in this region would be extremely unlikely. Thus the evidence in this family would also place the BRCA1 locus distal to tdj1474. However, the lower limit of this breakpoint is impossible to determine without parental genotype information. It is intriguing that the key recombinant breakpoint in this family confirms the result in Kindred 2082. As before, the localization information in this family is only meaningful if the breast cancer was due to the BRCA1 gene. However, her relatively early age at diagnosis (44) makes this seem very likely since the risk of breast cancer before age 45 in the general population is low (approximately 1%).

40 Kindred 2035

This family is similar to K1901 in that the information on the critical recombinant events is not directly observed but is inferred from the observation that the two haplotypes which are cosegregating with the early onset breast cancer in the two branches of the family appear identical for markers located in the proximal portion of the 17q BRCA1 region but differ at more distal loci. Each of these two haplotypes occurs in at least four cases of early-onset or bilateral breast cancer. The overall LOD score with ED2 in this family is 2.2, and considering that there is a case of ovarian cancer in the family (indicating a prior probability of BRCA1 linkage of 80%), the resulting posterior probability that this family is linked to BRCA1 is 0.998. The haplotypes are identical for the markers Mfd15, THRA1, Mfd191, ED2, AA1, D17S858 and D17S902. The common allele at Mfd15 and ED2 are both quite rare, indicating that this haplotype is shared identically by descent. The haplotypes are discordant, however, for CA375, 4-7, and Mfd188, and several more distal markers. This indicates that the BRCA1 locus must lie above the marker CA-375. This marker is located approximately 50 kb below D17S78, so it serves primarily as additional confirmation of this previous lower boundary as reported in Simard et al. (1993).

Kindred 1813

Kindred 1813 is a small family with four cases of breast cancer diagnosed at very early ages whose mother also had breast cancer diagnosed at an early age and ovarian cancer

5,753,441

49

some years later. This situation is somewhat complicated by the fact the four cases appear to have three different fathers, only one of whom has been genotyped. However, by typing a number of different markers in the BRCA1 region as well as highly polymorphic markers elsewhere in the genome, the paternity of all children in the family has been determined with a high degree of certainty. This family yields a maximum multipoint LOD score of 0.60 with 17q markers and, given that there is at least one case of ovarian cancer, results in a posterior probability of being a BRCA1 linked family of 0.93. This family contains a directly observable recombination event in individual 18 (see FIG. 5 in Simard et al., *Human Mol. Genet.* 2:1193-1199 (1993)), who developed breast cancer at age 34. The genotype of her affected mother at the relevant 17q loci can be inferred from her genotypes, her affected sister's genotypes, and the genotypes of three other unaffected siblings. Individual 18 inherits the BRCA1-linked alleles for the following loci: Mfd15, THRA1, D17S800, D17S855, AA1, and D17S931. However, for markers below D17S931, i.e., U5R, vrs31, D17S858, and D17S579, she has inherited the alleles located on the non-disease bearing chromosome. The evidence from this family therefore would place the BRCA1 locus proximal to the marker U5R. Because of her early age at diagnosis (34) it is extremely unlikely that the recombinant individual's cancer is not due to the gene responsible for the other cases of breast/ovarian cancer in this family; the uncertainty in this family comes from our somewhat smaller amount of evidence that breast cancer in this family is due to BRCA1 rather than a second, as yet unmapped, breast cancer susceptibility locus.

Size of the region containing BRCA1

Based on the genetic data described in detail above, the BRCA1 locus must lie in the interval between the markers tdj1474 and U5R, both of which were isolated in our laboratory. Based upon the physical maps shown in FIGS. 2 and 3, we can try to estimate the physical distance between these two loci. It takes approximately 14 P1 clones with an average insert size of approximately 80 kb to span the region. However, because all of these P1s overlap to some unknown degree, the physical region is most likely much smaller than 14 times 80 kb. Based on restriction maps of the clones covering the region, we estimate the size of the region containing BRCA1 to be approximately 650 kb.

EXAMPLE 7

Identification of Candidate cDNA Clones for the BRCA1 Locus by Genomic Analysis of the Contig Region

Complete screen of the plausible region.

The first method to identify candidate cDNAs, although labor intensive, used known techniques. The method comprised the screening of cosmids and P1 and BAC clones in the contig to identify putative coding sequences. The clones containing putative coding sequences were then used as probes on filters of cDNA libraries to identify candidate cDNA clones for future analysis. The clones were screened for putative coding sequences by either of two methods.

Zoo blots.

The first method for identifying putative coding sequences was by screening the cosmid and P1 clones for sequences conserved through evolution across several species. This technique is referred to as "zoo blot analysis" and is described by Monaco, 1986. Specifically, DNAs from cow, chicken, pig, mouse and rat were digested with the

50

restriction enzymes EcoRI and HindIII (8 µg of DNA per enzyme). The digested DNAs were separated overnight on an 0.7% gel at 20 volts for 16 hours (14 cm gel), and the DNA transferred to Nylon membranes using standard Southern blot techniques. For example, the zoo blot filter was treated at 65° C. in 0.1×SSC, 0.5% SDS, and 0.2M Tris, pH 8.0, for 30 minutes and then blocked overnight at 42° C. in 5× SSC, 10% PEG 8000, 20 mM NaPO₄ pH 6.8, 100 µg/ml Salmon Sperm DNA, 1× Denhardt's, 50% formamide, 0.1% SDS, and 2 µg/ml C₆t-1DNA.

The cosmid and P1 clones to be analyzed were digested with a restriction enzyme to release the human DNA from the vector DNA. The DNA was separated on a 14 cm, 0.5% agarose gel run overnight at 20 volts for 16 hours. The human DNA bands were cut out of the gel and electroeluted from the gel wedge at 100 volts for at least two hours in 0.5× Tris Acetate buffer (Maniatis et al., 1982). The eluted Not I digested DNA (~15 kb to 25 kb) was then digested with EcoRI restriction enzyme to give smaller fragments (~0.5 kb to 5.0 kb) which melt apart more easily for the next step of labeling the DNA with radionucleotides. The DNA fragments were labeled by means of the hexamer random prime labeling method (Boehringer-Mannheim, Cat. #1004760). The labeled DNA was spermine precipitated (add 100 µl TE, 5 µl 0.1M spermine, and 5 µl of 10 mg/ml salmon sperm DNA) to remove unincorporated radionucleotides. The labeled DNA was then resuspended in 100 µl TE, 0.5M NaCl at 65° C. for 5 minutes and then blocked with Human C₆t-1 DNA for 2-4 hrs. as per the manufacturer's instructions (Gibco/BRL, Cat. #5279SA). The C₆t-1 blocked probe was incubated on the zoo blot filters in the blocking solution overnight at 42° C. The filters were washed for 30 minutes at room temperature in 2×SSC, 0.1% SDS, and then in the same buffer for 30 minutes at 55° C. The filters were then exposed 1 to 3 days at -70° C. to Kodak XAR-5 film with an intensifying screen. Thus, the zoo blots were hybridized with either the pool of EcoRI fragments from the insert, or each of the fragments individually.

HTF island analysis.

The second method for identifying cosmids to use as probes on the cDNA libraries was HTF island analysis. Since the pulsed-field map can reveal HTF islands, cosmids that map to these HTF island regions were analyzed with priority. HTF islands are segments of DNA which contain a very high frequency of unmethylated CpG dinucleotides (Tonollio et al., 1990) and are revealed by the clustering of restriction sites of enzymes whose recognition sequences include CpG dinucleotides. Enzymes known to be useful in HTF-island analysis are AscI, NotI, BssHII, EagI, SacII, NaeI, NarI, SmaI, and MluI (Anand, 1992). A pulsed-field map was created using the enzymes NotI, NruI, EagI, SacII, and SalI, and two HTF islands were found. These islands are located in the distal end of the region, one being distal to the GP2B locus, and the other being proximal to the same locus, both outside the BRCA1 region. The cosmids derived from the YACs that cover these two locations were analyzed to identify those that contain these restriction sites, and thus the HTF islands.

cDNA screening.

Those clones that contain HTF islands or show hybridization to other species DNA besides human are likely to contain coding sequences. The human DNA from these clones was isolated as whole insert or as EcoRI fragments and labeled as described above. The labeled DNA was used to screen filters of various cDNA libraries under the same conditions as the zoo blots except that the cDNA filters

5,753,441

51

undergo a more stringent wash of $0.1\times$ SSC, 0.1% SDS at 65° C. for 30 minutes twice.

Most of the CDNA libraries used to date in our studies (libraries from normal breast tissue, breast tissue from a woman in her eighth month of pregnancy and a breast malignancy) were prepared at Clontech, Inc. The CDNA library generated from breast tissue of an 8 month pregnant woman is available from Clontech (Cat. #HL1037a) in the Lambda gt-10 vector, and is grown in C600OHf bacterial host cells. Normal breast tissue and malignant breast tissue samples were isolated from a 37 year old Caucasian female and one-gram of each tissue was sent to Clontech for mRNA processing and cDNA library construction. The latter two libraries were generated using both random and oligo-dT priming, with size selection of the final products which were then cloned into the Lambda Zap II vector, and grown in XL1-blue strain of bacteria as described by the manufacturer. Additional tissue-specific CDNA libraries include human fetal brain (Stratagene, Cat. 936206), human testis (Clontech Cat. HL3024), human thymus (Clontech Cat. HL1127n), human brain (Clontech Cat. HL11810), human placenta (Clontech Cat 1075b), and human skeletal muscle (Clontech Cat. HL 1124b).

The CDNA libraries were plated with their host cells on NZCYM plates, and filter lifts are made in duplicate from each plate as per Maniatis et al. (1982). Insert (human) DNA from the candidate genomic clones was purified and radioactively labeled to high specific activity. The radioactive DNA was then hybridized to the CDNA filters to identify those cDNAs which correspond to genes located within the candidate cosmid clone. cDNAs identified by this method were picked, replated, and screened again with the labeled clone insert or its derived EcoRI fragment DNA to verify their positive status. Clones that were positive after this second round of screening were then grown up and their DNA purified for Southern blot analysis and sequencing. Clones were either purified as plasmid through in vivo excision of the plasmid from the Lambda vector as described in the protocols from the manufacturers, or isolated from the Lambda vector as a restriction fragment and subcloned into plasmid vector.

The Southern blot analysis was performed in duplicate, one using the original genomic insert DNA as a probe to verify that cDNA insert contains hybridizing sequences. The second blot was hybridized with cDNA insert DNA from the largest cDNA clone to identify which clones represent the same gene. All cDNAs which hybridize with the genomic clone and are unique were sequenced and the DNA analyzed to determine if the sequences represent known or unique genes. All cDNA clones which appear to be unique were further analyzed as candidate BRCA1 loci. Specifically, the clones are hybridized to Northern blots to look for breast specific expression and differential expression in normal versus breast tumor RNAs. They are also analyzed by PCR on clones in the BRCA1 region to verify their location. To map the extent of the locus, full length cDNAs are isolated and their sequences used as PCR probes on the YACs and the clones surrounding and including the original identifying clones. Intron-exon boundaries are then further defined through sequence analysis.

We have screened the normal breast, 8 month pregnant breast and fetal brain cDNA libraries with zoo blot-positive Eco RI fragments from cosmid BAC and P1 clones in the region. Potential BRCA1 cDNA clones were identified among the three libraries. Clones were picked, replated, and screened again with the original probe to verify that they were positive.

52

Analysis of hybrid-selected cDNA.

cDNA fragments obtained from direct selection were checked by Southern blot hybridization against the probe DNA to verify that they originated from the contig. Those that passed this test were sequenced in their entirety. The set of DNA sequences obtained in this way were then checked against each other to find independent clones that overlapped. For example, the clones 694-65, 1240-1 and 1240-33 were obtained independently and subsequently shown to derive from the same contiguous cDNA sequence which has been named EST:489:1.

Analysis of candidate clones.

One or more of the candidate genes generated from above were sequenced and the information used for identification and classification of each expressed gene. The DNA sequences were compared to known genes by nucleotide sequence comparisons and by translation in all frames followed by a comparison with known amino acid sequences. This was accomplished using Genetic Data Environment (GDE) version 2.2 software and the Basic Local Alignment Search Tool (Blast) series of client/server software packages (e.g., BLASTN 1.3.13 MP), for sequence comparison against both local and remote sequence databases (e.g., GenBank), running on Sun SPARC workstations. Sequences reconstructed from collections of cDNA clones identified with the cosmids and P1s have been generated. All candidate genes that represented new sequences were analyzed further to test their candidacy for the putative BRCA1 locus.

Mutation screening.

To screen for mutations in the affected pedigrees, two different approaches were followed. First, genomic DNA isolated from family members known to carry the susceptibility allele of BRCA1 was used as a template for amplification of candidate gene sequences by PCR. If the PCR primers flank or overlap an intron/exon boundary, the amplified fragment will be larger than predicted from the cDNA sequence or will not be present in the amplified mixture. By a combination of such amplification experiments and sequencing of P1, BAC or cosmid clones using the set of designed primers it is possible to establish the intron/exon structure and ultimately obtain the DNA sequences of genomic DNA from the pedigrees.

A second approach that is much more rapid if the intron/exon structure of the candidate gene is complex involves sequencing fragments amplified from pedigree lymphocyte cDNA. cDNA synthesized from lymphocyte mRNA extracted from pedigree blood was used as a substrate for PCR amplification using the set of designed primers. If the candidate gene is expressed to a significant extent in lymphocytes, such experiments usually produce amplified fragments that can be sequenced directly without knowledge of intron/exon junctions.

The products of such sequencing reactions were analyzed by gel electrophoresis to determine positions in the sequence that contain either mutations such as deletions or insertions, or base pair substitutions that cause amino acid changes or other detrimental effects.

Any sequence within the BRCA1 region that is expressed in breast is considered to be a candidate gene for BRCA1. Compelling evidence that a given candidate gene corresponds to BRCA1 comes from a demonstration that pedigree families contain defective alleles of the candidate.

A000296

53
EXAMPLE 8

Identification of BRCA1

Identification of BRCA1.

Using several strategies, a detailed map of transcripts was developed for the 600 kb region of 17q21 between D17S1321 and D17S1324. Candidate expressed sequences were defined as DNA sequences obtained from: 1) direct screening of breast, fetal brain, or lymphocyte CDNA libraries, 2) hybrid selection of breast, lymphocyte or ovary cDNAs, or 3) random sequencing of genomic DNA and prediction of coding exons by XPOUND (Thomas and Skolnick, 1994). These expressed sequences in many cases were assembled into contigs composed of several independently identified sequences. Candidate genes may comprise more than one of these candidate expressed sequences. Sixty-five candidate expressed sequences within this region were identified by hybrid selection, by direct screening of CDNA libraries, and by random sequencing of P1 subclones. Expressed sequences were characterized by transcript size, DNA sequence, database comparison, expression pattern, genomic structure, and, most importantly, DNA sequence analysis in individuals from kindreds segregating 17q-linked breast and ovarian cancer susceptibility.

Three independent contigs of expressed sequence, 1141:1 (649 bp), 694:5 (213 bp) and 754:2 (1079 bp) were isolated and eventually shown to represent portions of BRCA1. When ESTs for these contigs were used as hybridization probes for Northern analysis, a single transcript of approximately 7.8 kb was observed in normal breast mRNA, suggesting that they encode different portions of a single gene. Screens of breast, fetal brain, thymus, testes, lymphocyte and placental cDNA libraries and PCR experiments with breast mRNA linked the 1141:1, 694:5 and 754:2 contigs. 5' RACE experiments with thymus, testes, and breast mRNA extended the contig to the putative 5' end, yielding a composite full length sequence. PCR and direct sequencing of P1s and BACs in the region were used to identify the location of introns and allowed the determination of splice donor and acceptor sites. These three expressed sequences were merged into a single transcription unit that proved in the final analysis to be BRCA1. This transcription unit is located adjacent to D17S855 in the center of the 600 kb region (FIG. 4).

Combination of sequences obtained from cDNA clones, hybrid selection sequences, and amplified PCR products allowed construction of a composite full length BRCA1 cDNA (SEQ ID NO:1). The sequence of the BRCA1 cDNA (up through the stop codon) has also been deposited with GenBank and assigned accession number U-14680. This deposited sequence is incorporated herein by reference. The cDNA clone extending farthest in the 3' direction contains a poly(A) tract preceded by a polyadenylation signal. Conceptual translation of the cDNA revealed a single long open reading frame of 208 kilodaltons (amino acid sequence: SEQ ID NO:2) with a potential initiation codon flanked by

sequences resembling the Kozak consensus sequence (Kozak, 1987). Smith-Waterman (Smith and Waterman, 1981) and BLAST (Altschul et al., 1990) searches identified a sequence near the amino terminus with considerable homology to zinc-finger domains (FIG. 5). This sequence contains cysteine and histidine residues present in the consensus C3HC4 zinc-finger motif and shares multiple other residues with zinc-finger proteins in the databases. The BRCA1 gene is composed of 23 coding exons arrayed over more than 100 kb of genomic DNA (FIG. 6). Northern blots using fragments of the BRCA1 cDNA as probes identified a single transcript of about 7.8 kb, present most abundantly in breast, thymus and testis, and also present in ovary (FIG. 7). Four alternatively spliced products were observed as independent cDNA clones; 3 of these were detected in breast and 2 in ovary mRNA (FIG. 6). A PCR survey from tissue cDNAs further supports the idea that there is considerable heterogeneity near the 5' end of transcripts from this gene; the molecular basis for the heterogeneity involves differential choice of the first splice donor site, and the changes detected all alter the transcript in the region 5' of the identified start codon. We have detected six potential alternate splice donors in this 5' untranslated region, with the longest deletion being 1,155 bp. The predominant form of the BRCA1 protein in breast and ovary lacks exon 4. The nucleotide sequence for BRCA1 exon 4 is shown in SEQ ID NO:11, with the predicted amino acid sequence shown in SEQ ID NO:12.

Additional 5' sequence of BRCA1 genomic DNA is set forth in SEQ ID NO:13. The G at position 1 represents the potential start site in testis. The A in position 140 represents the potential start site in somatic tissue. There are six alternative splice forms of this 5' sequence as shown in FIG. 8. The G at position 356 represents the canonical first splice donor site. The G at position 444 represents the first splice donor site in two clones (testis 1 and testis 2). The G at position 889 represents the first splice donor site in thymus 3. A fourth splice donor site is the G at position 1230. The T at position 1513 represents the splice acceptor site for all of the above splice donors. A fifth alternate splice form has a first splice donor site at position 349 with a first acceptor site at position 591 and a second splice donor site at position 889 and a second acceptor site at position 1513. A sixth alternate form is unspliced in this 5' region. The A at position 1532 is the canonical start site, which appears at position 120 of SEQ ID NO:1. Partial genomic DNA sequences determined for BRCA1 are set forth in FIGS. 10A-10H and SEQ ID Numbers:14-34. The lower case letters (in FIGS. 10A-10H) denote intron sequence while the upper case letters denote exon sequence. Indefinite intervals within introns are designated with vvvvvvvvvvvv in FIGS. 10A-10H. The intron/exon junctions are shown in Table 9. The CAG found at the 5' end of exons 8 and 14 is found in some cDNAs but not in others. Known polymorphic sites are shown in FIGS. 10A-10H in boldface type and are underlined.

TABLE 9

Exon	Base Position*			Intron Borders		
	No.	5'	3'	Length	3'	
e1		1	100	100	GATAAATTAAAACTGCGACTGCGCGGCGTG ^{35*}	GTAGTAGAGTCCCGGGAAGGGACAGGGGG ³⁶
e2		101	199	99	ATATATATATGTTTTTCTAATGIGITAAAG ³⁷	GTAAGTCAGCACAAAGAGTGTAITTAATTGG ³⁸
e3		200	253	54	TTTCTTTTCTCCCCCCCCCTACCTGCTAG ³⁹	GTAAGTTTGAATGTGTTAIGTGCGCTCCATT ⁴⁰

5,753,441

55

56

TABLE 9-continued

Exon				Intron Borders	
Base Position*					
No.	5'	3'	Length	5'	3'
e4	***	***	111	AGCTACTTTTTTTTTTTTTTTTIGAGACAG ⁴¹	GTAAGTGCACACCACCATATCCAGCTAAAT ⁴²
e5	254	331	78	AATTGTTCTTTCTTTCTTTATAATTATAG ⁴³	GTAATAAATTTGGTAATGATGCTAGGTTGG ⁴⁴
e6	332	420	89	GAGTGTGTTCTCAACAATTAAATTCAG ⁴⁵	GTAAGTGTGAATAATCCCAAGAATGACACT ⁴⁶
e7	421	560	140	AAACATAATGTTTTCCCTTGTAITTTACAG ⁴⁷	GTAATAACCATTTGTTTTCTTCTTCTTCTC ⁴⁸
e8	561	666	106	TGCTTGACTGTCTTTTACCATACGTITTAG ⁴⁹	GTAAGGCTCTCAGGTTTTTTAAGTATTAA ⁵⁰
e9	667	712	46	TGATTATTTTTTGGGGGAAATTTTAAAG ⁵¹	GTGAGTCAAAAGAGAACCTTTGTCTATGAAG ⁵²
e10	713	789	77	TCTTATAGGACTCTGTCTTTTCCCTATAG ⁵³	GTAATGGCAAAGTTTGCCAACCTTAACAGGC ⁵⁴
e11	790	4215	3426	GAGTACCTTGTTATTTTGTATATTTTCAG ⁵⁵	GTAATTGGAACAGGTTTTTGTGTTTGCCCC ⁵⁶
e12	4216	4302	87	ACAICTGAACCTCTGTTTTTGTATTAAAG ⁵⁷	AGGTAAAAAGCGTGTGTGTGTGCACATG ⁵⁸
e13	4303	4476	174	CAITTTCTTGGTACCAITTAATCATTITTA ⁵⁹	GTGTGTATTGTTGGCCAAACACTGATATCT ⁶⁰
e14	4477	4603	127	AGTAGATTGTTTTCTCATTCATTTAAAG ⁶¹	GTAAGAAACATCAATGAAAAGATGCTGTGG ⁶²
e15	4604	4794	191	ATGGTTTTCTCTCCATTATCTTCTAG ^{63**}	GTAATATTTCATCTGCTGATTTGGAACAAA ⁶⁴
e16	4795	5105	311	TGTAAATAACTTCTCCCATTCCTTTCAG ⁶⁵	GTGAGTGATCCATATGATCTCCCTAATG ⁶⁶
e17	5106	5193	88	ATGATAATGGAATATTGATTAAATTCAG ⁶⁷	GTATACCAAGAACCTTTACAGAATACCTTGG ⁶⁸
e18	5194	5271	78	CTAATCCITTGAGTGTTTTCAITCTGCAG ⁶⁹	GTAAGTATAATACATTTCTCCCTCTCTCC ⁷⁰
e19	5272	5312	41	TGTAACTGTCTTTTCTATGATCTCTTAA ⁷¹	GTAAGTACTTGATGTTACAACTAACACAGA ⁷²
e20	5313	5396	84	TCCTGATGGGTGTGTGTTGTTCTTTCAG ⁷³	GTAAGCTCCCTCCCTCAAGTGTACAAAG ⁷⁴
e21	5397	5451	55	CTGCCCTCTCTCTCTCTCTCTCTTCCAG ⁷⁵	GTAAGAGCCTGGGAGAACCCAGAGTTCCA ⁷⁶
e22	5452	5525	74	AGTGATTTTACATGAAATGTCCATTTAG ⁷⁷	GTAAGTATTGGGTGCCCTGTCTAGTGTGGGA ⁷⁸
e23	5526	5586	61	TGAAATGCTCTTCTCTCTCTGGGGATCCAG ⁷⁹	GTAAGGTGCTCGCATGTACCTGTGCTATT ⁸⁰
e24	5587	5914	328	CTAATCTCTGCTGTGTCTCTGTCTCCAG ⁸¹	

*Base numbers in SEQ ID NO: 1.
**Numbers in superscript refer to SEQ ID NOS.
***e4 from SEQ ID NO: 11.

Low stringency blots in which genomic DNA from organisms of diverse phylogenetic background were probed with BRCA1 sequences that lack the zinc-finger region revealed strongly hybridizing fragments in human, monkey, sheep and pig, and very weak hybridization signals in rodents. This result indicates that, apart from the zinc-finger domain, BRCA1 is conserved only at a moderate level through evolution.

Germ-line BRCA1 mutations in 17q-linked kindreds. The most rigorous test for BRCA1 candidate genes is to search for potentially disruptive mutations in carrier individuals from kindreds that segregate 17q-linked susceptibility to breast and ovarian cancer. Such individuals must contain BRCA1 alleles that differ from the wildtype sequence. The set of DNA samples used in this analysis consisted of DNA from individuals representing 8 different BRCA1 kindreds (Table 10).

TABLE 10

KINDRED DESCRIPTIONS AND ASSOCIATED LOD SCORES						
Kindred	Cases (n)			Sporadic Cases ¹ (n)	Score	Marker(s)
	Br	Br < 50	Ov			
2082	31	20	22	7	9.49	D17S1327
2099	22	14	2*	0	2.36	D17S800/D17S855 ²
2035	10	8	1*	0	2.25	D17S1327
1901	10	7	1*	0	1.50	D17S855
1925	4	3	0	0	0.55	D17S579
1910	5	4	0	0	0.36	D17S579/D17S250 ²
1927	5	4	0	1	-0.44	D17S250
1911	8	5	0	2	-0.20	D17S250

¹Number of women with breast cancer (diagnosed under age 50) or ovarian cancer (diagnosed at any age) who do not share the BRCA1-linked haplotype segregating in the remainder of the cases in the kindred.
²Multipoint LOD score calculated using both markers
*kindred contains one individual who had both breast and ovarian cancer; this individual is counted as a breast cancer case and as an ovarian cancer case.

The logarithm of the odds (LOD) scores in these kindreds range from 9.49 to -0.44 for a set of markers in 17q21. Four

of the families have convincing LOD scores for linkage, and 4 have low positive or negative LOD scores. The latter kindreds were included because they demonstrate haplotype sharing at chromosome 17q21 for at least 3 affected members. Furthermore, all kindreds in the set display early age of breast cancer onset and 4 of the kindreds include at least one case of ovarian cancer, both hallmarks of BRCA1 kindreds. One kindred, 2082, has nearly equal incidence of breast and ovarian cancer, an unusual occurrence given the relative rarity of ovarian cancer in the population. All of the kindreds except two were ascertained in Utah. K2035 is from the midwest. K2099 is an African-American kindred from the southern U.S.A.

In the initial screen for predisposing mutations in BRCA1, DNA from one individual who carries the predisposing haplotype in each kindred was tested. The 23 coding exons and associated splice junctions were amplified either from genomic DNA samples or from CDNA prepared from lymphocyte mRNA. When the amplified DNA sequences were compared to the wildtype sequence, 4 of the 8 kindred samples were found to contain sequence variants (Table 11).

TABLE 11

PREDISPOSING MUTATIONS			
Kindred Number	Mutation	Coding Effect	Location*
2082	C→T	Gln→Stop	4056
1910	extra C	frameshift	5385
2099	T→G	Met→Arg	5443
2035	?	loss of transcript	
1901	11 bp deletion	frameshift	189

*In Sequence ID NO: 1

All four sequence variants are heterozygous and each appears in only one of the kindreds. Kindred 2082 contains a nonsense mutation in exon 11 (FIG. 9A), Kindred 1910 contains a single nucleotide insertion in exon 20 (FIG. 9B), and Kindred 2099 contains a missense mutation in exon 21, resulting in a Met→Arg substitution. The frameshift and

nonsense mutations are likely disruptive to the function of the BRCA1 product. The peptide encoded by the frameshift allele in Kindred 1910 would contain an altered amino acid sequence beginning 108 residues from the wildtype C-terminus. The peptide encoded by the frameshift allele in Kindred 1901 would contain an altered amino acid sequence beginning with the 24th residue from the wildtype N-terminus. The mutant allele in Kindred 2082 would encode a protein missing 551 residues from the C-terminus. The missense substitution observed in Kindred 2099 is potentially disruptive as it causes the replacement of a small hydrophobic amino acid (Met), by a large charged residue (Arg). Eleven common polymorphisms were also identified, 8 in coding sequence and 3 in introns.

The individual studied in Kindred 2035 evidently contains a regulatory mutation in BRCA1. In her cDNA, a polymorphic site (A→G at base 3667) appeared homozygous, whereas her genomic DNA revealed heterozygosity at this position (FIG. 9C). A possible explanation for this observation is that mRNA from her mutated BRCA1 allele is absent due to a mutation that affects its production or stability. This possibility was explored further by examining 5 polymorphic sites in the BRCA1 coding region, which are separated by as much as 3.5 kb in the BRCA1 transcript. In all cases where her genomic DNA appeared heterozygous for a polymorphism, cDNA appeared homozygous. In individuals from other kindreds and in non-haplotype carriers in Kindred 2035, these polymorphic sites could be observed as heterozygous in cDNA, implying that amplification from cDNA was not biased in favor of one allele. This analysis indicates that a BRCA1 mutation in Kindred 2035 either prevents transcription or causes instability or aberrant splicing of the BRCA1 transcript.

Cosegregation of BRCA1 mutations with BRCA1 haplotypes and population frequency analysis.

In addition to potentially disrupting protein function, two criteria must be met for a sequence variant to qualify as a candidate predisposing mutation. The variant must: 1) be present in individuals from the kindred who carry the predisposing BRCA1 haplotype and absent in other members of the kindred, and 2) be rare in the general population.

Each mutation was tested for cosegregation with BRCA1. For the frameshift mutation in Kindred 1910, two other haplotype carriers and one non-carrier were sequenced (FIG. 9B). Only the carriers exhibited the frameshift mutation. The C to T change in Kindred 2082 created a new AvrII restriction site. Other carriers and non-carriers in the kindred were tested for the presence of the restriction site (FIG. 9A). An allele-specific oligonucleotide (ASO) was designed to detect the presence of the sequence variant in Kindred 2099. Several individuals from the kindred, some known to carry the haplotype associated with the predisposing allele, and others known not to carry the associated haplotype, were screened by ASO for the mutation previously detected in the kindred. In each kindred, the corresponding mutant allele was detected in individuals carrying the BRCA1-associated haplotype, and was not detected in noncarriers. In the case of the potential regulatory mutation observed in the individual from Kindred 2035, cDNA and genomic DNA from carriers in the kindred were compared for heterozygosity at polymorphic sites. In every instance, the extinguished allele in the cDNA sample was shown to lie on the chromosome that carries the BRCA1 predisposing allele (FIG. 9C).

To exclude the possibility that the mutations were simply common polymorphisms in the population, ASOs for each mutation were used to screen a set of normal DNA samples.

Gene frequency estimates in Caucasians were based on random samples from the Utah population. Gene frequency estimates in African-Americans were based on 39 samples provided by M. Peracek-Vance which originate from African-Americans used in her linkage studies and 20 newborn Utah African-Americans. None of the 4 potential predisposing mutations was found in the appropriate control population, indicating that they are rare in the general population. Thus, two important requirements for BRCA1 susceptibility alleles were fulfilled by the candidate predisposing mutations: 1) cosegregation of the mutant allele with disease, and 2) absence of the mutant allele in controls, indicating a low gene frequency in the general population.

Phenotypic Expression of BRCA1 Mutations.

The effect of the mutations on the BRCA1 protein correlated with differences in the observed phenotypic expression in the BRCA1 kindreds. Most BRCA1 kindreds have a moderately increased ovarian cancer risk, and a smaller subset have high risks of ovarian cancer, comparable to those for breast cancer (Easton et al., 1993). Three of the four kindreds in which BRCA1 mutations were detected fall into the former category, while the fourth (K2082) falls into the high ovarian cancer risk group. Since the BRCA1 nonsense mutation found in K2082 lies closer to the amino terminus than the other mutations detected, it might be expected to have a different phenotype. In fact, Kindred K2082 mutation has a high incidence of ovarian cancer, and a later mean age at diagnosis of breast cancer cases than the other kindreds (Goldgar et al., 1994). This difference in age of onset could be due to an ascertainment bias in the smaller, more highly penetrant families, or it could reflect tissue-specific differences in the behavior of BRCA1 mutations. The other 3 kindreds that segregate known BRCA1 mutations have, on average, one ovarian cancer for every 10 cases of breast cancer, but have a high proportion of breast cancer cases diagnosed in their late 20's or early 30's. Kindred 1910, which has a frameshift mutation, is noteworthy because three of the four affected individuals had bilateral breast cancer, and in each case the second tumor was diagnosed within a year of the first occurrence. Kindred 2035, which segregates a potential regulatory BRCA1 mutation, might also be expected to have a dramatic phenotype. Eighty percent of breast cancer cases in this kindred occur under age 50. This figure is as high as any in the set, suggesting a BRCA1 mutant allele of high penetrance (Table 10).

Although the mutations described above clearly are deleterious, causing breast cancer in women at very young ages, each of the four kindreds with mutations includes at least one woman who carries the mutation who lived until age 80 without developing a malignancy. It will be of utmost importance in the studies that follow to identify other genetic or environmental factors that may ameliorate the effects of BRCA1 mutations.

In four of the eight putative BRCA1-linked kindreds, potential predisposing mutations were not found. Three of these four have LOD scores for BRCA1-linked markers of less than 0.55. Thus, these kindreds may not in reality segregate BRCA1 predisposing alleles. Alternatively, the mutations in these four kindreds may lie in regions of BRCA1 that, for example, affect the level of transcript and therefore have thus far escaped detection.

Role of BRCA1 in Cancer.

Most tumor suppressor genes identified to date give rise to protein products that are absent, nonfunctional, or reduced in function. The majority of TP53 mutations are missense;

some of these have been shown to produce abnormal p53 molecules that interfere with the function of the wildtype product (Shaullian et al., 1992; Srivastava et al., 1993). A similar dominant negative mechanism of action has been proposed for some adenomatous polyposis coli (APC) alleles that produce truncated molecules (Su et al., 1993), and for point mutations in the Wilms' tumor gene (WT1) that alter DNA binding of the protein (Little et al., 1993). The nature of the mutations observed in the BRCA1 coding sequence is consistent with production of either dominant negative proteins or nonfunctional proteins. The regulatory mutation inferred in Kindred 2035 cannot be a dominant negative; rather, this mutation likely causes reduction or complete loss of BRCA1 expression from the affected allele.

The BRCA1 protein contains a C₃HC₄ zinc-finger domain, similar to those found in numerous DNA binding proteins and implicated in zinc-dependent binding to nucleic acids. The first 180 amino acids of BRCA1 contain five more basic residues than acidic residues. In contrast, the remainder of the molecule is very acidic, with a net excess of 70 acidic residues. The excess negative charge is particularly concentrated near the C-terminus. Thus, one possibility is that BRCA1 encodes a transcription factor with an N-terminal DNA binding domain and a C-terminal transactivational "acidic blob" domain. Interestingly, another familial tumor suppressor gene, WT1, also contains a zinc-finger motif (Haber et al., 1990). Many cancer predisposing mutations in WT1 alter zinc-finger domains (Little et al., 1993; Haber et al., 1990; Little et al., 1992). WT1 encodes a transcription factor, and alternative splicing of exons that encode parts of the zinc-finger domain alter the DNA binding properties of WT1 (Bickmore et al., 1992). Some alternatively spliced forms of WT1 mRNA generate molecules that act as transcriptional repressors (Drummond et al., 1994). Some BRCA1 splicing variants may alter the zinc-finger motif, raising the possibility that a regulatory mechanism similar to that which occurs in WT1 may apply to BRCA1.

EXAMPLE 9

Analysis of Tumors for BRCA1 Mutations

To focus the analysis on tumors most likely to contain BRCA1 mutations, primary breast and ovarian carcinomas were typed for LOH in the BRCA1 region. Three highly polymorphic, simple tandem repeat markers were used to assess LOH: D17S1323 and D17S855, which are intragenic to BRCA1, and D17S1327, which lies approximately 100 kb distal to BRCA1. The combined LOH frequency in informative cases (i.e., where the germline was heterozygous) was 32/72 (44%) for the breast carcinomas and 12/21 (57%) for the ovarian carcinomas, consistent with previous measurements of LOH in the region (Futreal et al., 1992b; Jacobs et al., 1993; Sato et al., 1990; Eccles et al., 1990; Cropp et al., 1994). The analysis thus defined a panel of 32 breast tumors and 12 ovarian tumors of mixed race and age of onset to be examined for BRCA mutations. The complete 5,589 bp coding region and intron/exon boundary sequences of the gene were screened in this tumor set by direct sequencing alone or by a combination of single-strand conformation analysis (SSCA) and direct sequencing.

A total of six mutations (of which two are identical) was found, one in an ovarian tumor, four in breast tumors and one in a male unaffected haplotype carrier (Table 12). One mutation, Glu1541Ter, introduced a stop codon that would create a truncated protein missing 323 amino acids at the carboxy terminus. In addition, two missense mutations were

identified. These are Ala1708Glu and Met1775Arg and involve substitutions of small, hydrophobic residues by charged residues. Patients 17764 and 19964 are from the same family. In patient OV24 nucleotide 2575 is deleted and in patients 17764 and 19964 nucleotides 2993-2996 are deleted.

TABLE 12

Predisposing Mutations					
Patient	Codon	Nucleotide Change	Amino Acid Change	Age of Onset	Family History
BT098	1541	GAG → TAG	Glu → Stop	39	-
OV24	819	1 bp deletion	frameshift	44	-
BT106	1708	GCG → GAG	Ala → Glu	24	+
MC44	1775	ATG → AGG	Met → Arg	42	+
17764	958	4 bp deletion	frameshift	31	+
19964	958	4 bp deletion	frameshift		+

*Unaffected haplotype carrier, male

Several lines of evidence suggest that all five mutations represent BRCA1 susceptibility alleles:

- (i) all mutations are present in the germline;
- (ii) all are absent in appropriate control populations, suggesting they are not common polymorphisms;
- (iii) each mutant allele is retained in the tumor, as is the case in tumors from patients belonging to kindreds that segregate BRCA1 susceptibility alleles (Smith et al., 1992; Kelsell et al., 1993) (if the mutations represented neutral polymorphisms, they should be retained in only 50% of the cases);
- (iv) the age of onset in the four breast cancer cases with mutations varied between 24 and 42 years of age, consistent with the early age of onset of breast cancer in individuals with BRCA1 susceptibility; similarly, the ovarian cancer case was diagnosed at 44, an age that falls in the youngest 13% of all ovarian cancer cases; and finally,
- (v) three of the five cases have positive family histories of breast or ovarian cancer found retrospectively in their medical records, although the tumor set was not selected with regard to this criterion.

BT106 was diagnosed at a very early age with breast cancer. Her mother had ovarian cancer, her father had melanoma, and her paternal grandmother also had breast cancer. Patient MC44, an African-American, had bilateral breast cancer at an early age. This patient had a sister who died of breast cancer at a very early age, another sister who died of lymphoma, and a brother who died of lung cancer. Her mutation (Met1775Arg) had been detected previously in Kindred 2099, an African-American family that segregates a BRCA1 susceptibility allele, and was absent in African-American and Caucasian controls. Patient MC44, to our knowledge, is unrelated to Kindred 2099. The detection of a rare mutant allele, once in a BRCA1 kindred and once in the germline of an apparently unrelated early-onset breast cancer case, suggests that the Met1775Arg change may be a common predisposing mutation in African-Americans. Collectively, these observations indicate that all four BRCA1 mutations in tumors represent susceptibility alleles; no somatic mutations were detected in the samples analyzed.

The paucity of somatic BRCA1 mutations is unexpected, given the frequency of LOH on 17q, and the usual role of susceptibility genes as tumor suppressors in cancer progression. There are three possible explanations for this result: (i) some BRCA1 mutations in coding sequences were missed

5,753,441

61

by our screening procedure; (ii) BRCA1 somatic mutations fall primarily outside the coding exons; and (iii) LOH events in 17q do not reflect BRCA1 somatic mutations.

If somatic BRCA1 mutations truly are rare in breast and ovary carcinomas, this would have strong implications for the biology of BRCA1. The apparent lack of somatic BRCA1 mutations implies that there may be some fundamental difference in the genesis of tumors in genetically predisposed BRCA1 carriers, compared with tumors in the general population. For example, mutations in BRCA1 may have an effect only on tumor formation at a specific stage early in breast and ovarian development. This possibility is consistent with a primary function for BRCA1 in premenopausal breast cancer. Such a model for the role of BRCA1 in breast and ovarian cancer predicts an interaction between reproductive hormones and BRCA1 function. However, no clinical or pathological differences in familial versus sporadic breast and ovary tumors, other than age of onset, have been described (Lynch et al., 1990). On the other hand, the recent finding of increased TP53 mutation and microsatellite instability in breast tumors from patients with a family history of breast cancer (Glebov et al., 1994) may reflect some difference in tumors that arise in genetically predisposed persons. The involvement of BRCA1 in this phenomenon can now be addressed directly. Alternatively, the lack of somatic BRCA1 mutations may result from the existence of multiple genes that function in the same pathway of tumor suppression as BRCA1, but which collectively represent a more favored target for mutation in sporadic tumors. Since mutation of a single element in a genetic pathway is generally sufficient to disrupt the pathway, BRCA1 might mutate at a rate that is far lower than the sum of the mutational rates of the other elements.

EXAMPLE 10

Analysis of the BRCA1 Gene

The structure and function of BRCA1 gene are determined according to the following methods.

Biological Studies.

Mammalian expression vectors containing BRCA1 cDNA are constructed and transfected into appropriate breast carcinoma cells with lesions in the gene. Wild-type BRCA1 cDNA as well as altered BRCA1 cDNA are utilized. The altered BRCA1 cDNA can be obtained from altered BRCA1 alleles or produced as described below. Phenotypic reversion in cultures (e.g., cell morphology, doubling time, anchorage-independent growth) and in animals (e.g., tumorigenicity) is examined. The studies will employ both wild-type and mutant forms (Section B) of the gene.

Molecular Genetics Studies.

In vitro mutagenesis is performed to construct deletion mutants and missense mutants (by single base-pair substitutions in individual codons and cluster charged→alanine scanning mutagenesis). The mutants are used in biological, biochemical and biophysical studies.

Mechanism Studies.

The ability of BRCA1 protein to bind to known and unknown DNA sequences is examined. Its ability to trans-activate promoters is analyzed by transient reporter expression systems in mammalian cells. Conventional procedures such as particle-capture and yeast two-hybrid system are used to discover and identify any functional partners. The nature and functions of the partners are characterized. These partners in turn are targets for drug discovery.

62

Structural Studies.

Recombinant proteins are produced in *E. coli*, yeast, insect and/or mammalian cells and are used in crystallographical and NMR studies. Molecular modeling of the proteins is also employed. These studies facilitate structure-driven drug design.

EXAMPLE 11

Two Step Assay to Detect the Presence of BRCA1 in a Sample

Patient sample is processed according to the method disclosed by Antonarakis et al. (1985), separated through a 1% agarose gel and transferred to nylon membrane for Southern blot analysis.

Membranes are UV cross linked at 150 mJ using a GS Gene Linker (Bio-Rad). BRCA1 probe corresponding to nucleotide positions 3631–3930 of SEQ ID NO:1 is subcloned into pTZ18U. The phagemids are transformed into *E. coli* MV1190 infected with M13KO7 helper phage (Bio-Rad, Richmond, Calif.). Single stranded DNA is isolated according to standard procedures (see Sambrook et al., 1989).

Blots are prehybridized for 15–30 min at 65° C. in 7% sodium dodecyl sulfate (SDS) in 0.5M NaPO₄. The methods follow those described by Nguyen et al., 1992. The blots are hybridized overnight at 65° C. in 7% SDS, 0.5M NaPO₄ with 25–50 ng/ml single stranded probe DNA. Post-hybridization washes consist of two 30 min washes in 5% SDS, 40 mM NaPO₄ at 65° C., followed by two 30 min washes in 1% SDS, 40 mM NaPO₄ at 65° C.

Next the blots are rinsed with phosphate buffered saline (pH 6.8) for 5 min at room temperature and incubated with 0.2% casein in PBS for 30–60 min at room temperature and rinsed in PBS for 5 min. The blots are then preincubated for 5–10 minutes in a shaking water bath at 45° C. with hybridization buffer consisting of 6M urea, 0.3M NaCl, and 5× Denhardt's solution (see Sambrook, et al., 1989). The buffer is removed and replaced with 50–75 µl/cm² fresh hybridization buffer plus 2.5 nM of the covalently cross-linked oligonucleotide-alkaline phosphatase conjugate with the nucleotide sequence complementary to the universal primer site (UP-AP, Bio-Rad). The blots are hybridized for 20–30 min at 45° C. and post hybridization washes are incubated at 45° C. as two 10 min washes in 6M urea, 1× standard saline citrate (SSC), 0.1% SDS and one 10 min wash in 1× SSC, 0.1% TRITON®X-100, detergent. The blots are rinsed for 10 min at room temperature with 1× SSC.

Blots are incubated for 10 min at room temperature with shaking in the substrate buffer consisting of 0.1M diethanolamine, 1 mM MgCl₂, 0.02% sodium azide, pH 10.0. Individual blots are placed in heat sealable bags with substrate buffer and 0.2 mM AMPPD (3-(2'-spiroadamantane)-4-methoxy-4-(3'-phosphoryloxy)phenyl-1,2-dioxetane, disodium salt, Bio-Rad). After a 20 min incubation at room temperature with shaking, the excess AMPPD solution is removed. The blot is exposed to X-ray film overnight. Positive bands indicate the presence of BRCA1.

EXAMPLE 12

Generation of Polyclonal Antibody against BRCA1

Segments of BRCA1 coding sequence were expressed as fusion protein in *E. coli*. The overexpressed protein was purified by gel elution and used to immunize rabbits and mice using a procedure similar to the one described by

5,753,441

63

Harlow and Lane, 1988. This procedure has been shown to generate Abs against various other proteins (for example, see Kraemer et al., 1993).

Briefly, a stretch of BRCA1 coding sequence was cloned as a fusion protein in plasmid PET5A (Novagen, Inc., Madison, Wis.). The BRCA1 incorporated sequence includes the amino acids corresponding to #1361–1554 of SEQ ID NO:2. After induction with IPTG, the overexpression of a fusion protein with the expected molecular weight was verified by SDS/PAGE. Fusion protein was purified from the gel by electroelution. The identification of the protein as the BRCA1 fusion product was verified by protein sequencing at the N-terminus. Next, the purified protein was used as immunogen in rabbits. Rabbits were immunized with 100 µg of the protein in complete Freund's adjuvant and boosted twice in 3 week intervals, first with 100 µg of immunogen in incomplete Freund's adjuvant followed by 100 µg of immunogen in PBS. Antibody containing serum is collected two weeks thereafter.

This procedure is repeated to generate antibodies against the mutant forms of the BRCA1 gene. These antibodies, in conjunction with antibodies to wild type BRCA1, are used to detect the presence and the relative level of the mutant forms in various tissues and biological fluids.

EXAMPLE 13

Generation of Monoclonal Antibodies Specific for BRCA1

Monoclonal antibodies are generated according to the following protocol. Mice are immunized with immunogen comprising intact BRCA1 or BRCA1 peptides (wild type or mutant) conjugated to keyhole limpet hemocyanin using glutaraldehyde or EDC as is well known.

The immunogen is mixed with an adjuvant. Each mouse receives four injections of 10 to 100 µg of immunogen and after the fourth injection blood samples are taken from the mice to determine if the serum contains antibody to the immunogen. Serum titer is determined by ELISA or RIA. Mice with sera indicating the presence of antibody to the immunogen are selected for hybridoma production.

Spleens are removed from immune mice and a single cell suspension is prepared (see Harlow and Lane, 1988). Cell fusions are performed essentially as described by Kohler and Milstein, 1975. Briefly, P3.65.3 myeloma cells (American Type Culture Collection, Rockville, Md.) are fused with immune spleen cells using polyethylene glycol as described by Harlow and Lane, 1988. Cells are plated at a density of 2×10^5 cells/well in 96 well tissue culture plates. Individual wells are examined for growth and the supernatants of wells with growth are tested for the presence of BRCA1 specific antibodies by ELISA or RIA using wild type or mutant BRCA1 target protein. Cells in positive wells are expanded and subcloned to establish and confirm monoclonality.

Clones with the desired specificities are expanded and grown as ascites in mice or in a hollow fiber system to produce sufficient quantities of antibody for characterization and assay development.

EXAMPLE 14

Sandwich Assay for BRCA1

Monoclonal antibody is attached to a solid surface such as a plate, tube, bead, or particle. Preferably, the antibody is attached to the well surface of a 96-well ELISA plate. 100 μ l sample (e.g., serum, urine, tissue cytosol) containing the

64

BRCA1 peptide/protein (wild-type or mutant) is added to the solid phase antibody. The sample is incubated for 2 hrs at room temperature. Next the sample fluid is decanted, and the solid phase is washed with buffer to remove unbound material. 100 μ l of a second monoclonal antibody (to a different determinant on the BRCA1 peptide/protein) is added to the solid phase. This antibody is labeled with a detector molecule (e.g., 125 I, enzyme, fluorophore, or a chromophore) and the solid phase with the second antibody is incubated for two hrs at room temperature. The second antibody is decanted and the solid phase is washed with buffer to remove unbound material.

The amount of bound label, which is proportional to the amount of BRCA1 peptide/protein present in the sample, is quantitated. Separate assays are performed using monoclonal antibodies which are specific for the wild-type BRCA1 as well as monoclonal antibodies specific for each of the mutations identified in BRCA1.

Industrial Utility

As previously described above, the present invention provides materials and methods for use in testing BRCA1 alleles of an individual and an interpretation of the normal or predisposing nature of the alleles. Individuals at higher than normal risk might modify their lifestyles appropriately. In the case of BRCA1, the most significant non-genetic risk factor is the protective effect of an early, full term pregnancy. Therefore, women at risk could consider early childbearing or a therapy designed to simulate the hormonal effects of an early full-term pregnancy. Women at high risk would also strive for early detection and would be more highly motivated to learn and practice breast self examination. Such women would also be highly motivated to have regular mammograms, perhaps starting at an earlier age than the general population. Ovarian screening could also be undertaken at greater frequency. Diagnostic methods based on sequence analysis of the BRCA1 locus could also be applied to tumor detection and classification. Sequence analysis could be used to diagnose precursor lesions. With the evolution of the method and the accumulation of information about BRCA1 and other causative loci, it could become possible to separate cancers into benign and malignant.

Women with breast cancers may follow different surgical procedures if they are predisposed, and therefore likely to have additional cancers, than if they are not predisposed. Other therapies may be developed, using either peptides or small molecules (rational drug design). Peptides could be the missing gene product itself or a portion of the missing gene product. Alternatively, the therapeutic agent could be another molecule that mimics the deleterious gene's function, either a peptide or a nonpeptidic molecule that seeks to counteract the deleterious effect of the inherited locus. The therapy could also be gene based, through introduction of a normal BRCA1 allele into individuals to make a protein which will counteract the effect of the deleterious allele. These gene therapies may take many forms and may be directed either toward preventing the tumor from forming, curing a cancer once it has occurred, or stopping a cancer from metastasizing.

It will be appreciated that the methods and compositions of the instant invention can be incorporated in the form of a variety of embodiments, only a few of which are disclosed herein. It will be apparent to the artisan that other embodiments exist and do not depart from the spirit of the invention. Thus, the described embodiments are illustrative and should not be construed as restrictive.

LIST OF REFERENCES

Altschul, S. F. et al. (1990). *J. Mol. Biol.* 215: 195–197.

5,753,441

65

66

- American Cancer Society, Cancer Facts & Figures—1992. (American Cancer Society, Atlanta, Ga.).
- Anand, R. (1992). *Techniques for the Analysis of Complex Genomes*. (Academic Press).
- Anderson, et al. (1980). *Proc. Natl. Acad. Sci. USA* 77:5399–5403.
- Anderson, D. E. (1972). *J. Natl. Cancer Inst.* 48:1029–1034.
- Anderson, J. A., et al. (1992). *J. Otolaryngology* 21:321.
- Antonarakis, S. E., et al. (1985). *New Eng J. Med.* 313:842–848.
- Ausubel, F. M., et al. (1992). *Current Protocols in Molecular Biology*. (J. Wiley and Sons, N.Y.)
- Beaucage & Carruthers (1981). *Tetra. Letts.* 22:1859–1862.
- Berkner (1992). *Curr. Top. Microbiol. Immunol.* 158:39–61.
- Berkner, et al. (1988). *BioTechniques* 6:616–629.
- Bickmore, W. A., et al. (1992). *Science* 257:235–7.
- Bishop, D. T., et al. (1988). *Genet. Epidemiol.* 5:151–169.
- Bishop, D. T. and Gardner, E. J. (1980). In: *Banbury Report 4: Cancer Incidence in Defined Populations* (J. Cairns, J. L. Lyon, M. Skolnick, eds.), Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 309–408.
- Botstein, et al. (1980). *Am. J. Hum. Genet.* 32:314–331.
- Bowcock, A. M., et al. (1993). *Am. J. Hum. Genet.* 52:718.
- Bradyopadhyay and Temin (1984). *Mol. Cell. Biol.* 4:749–754.
- Breakfield and Geller (1987). *Mol. Neurobiol.* 1:337–371.
- Brinster, et al. (1981). *Cell* 27:223–231.
- Buchschacher and Panganiban (1992). *J. Virol.* 66:2731–2739.
- Buckler, et al. (1991). *Proc. Natl. Acad. Sci. USA* 88:4005–4009.
- Cannon-Albright, L., et al. (1994). *Cancer Research* 54:2378–2385.
- Capecchi, M. R. (1989). *Science* 244:1288.
- Cariello (1988). *Human Genetics* 42:726.
- Claus, E., et al. (1991). *Am. J. Hum. Genet.* 48:232–242.
- Conner, B. J., et al. (1983). *Proc. Natl. Acad. Sci. USA* 80:278–282.
- Constantini and Lacy (1981). *Nature* 294:92–94.
- Cotten, et al. (1990). *Proc. Natl. Acad. Sci. USA* 87:4033–4037.
- Cotton, et al. (1988). *Proc. Natl. Acad. Sci. USA* 85:4397.
- Cropp, C. S., et al. (1994). *Cancer Res.* 54:2548–2551.
- Culver, et al. (1992). *Science* 256:1550–1552.
- Curiel, et al. (1991a). *Proc. Natl. Acad. Sci. USA* 88:8850–8854.
- Curiel, et al. (1991b). *Hum. Gene Ther.* 3:147–154.
- Deutscher, M. (1990). *Meth. Enzymology* 182 (Academic Press, San Diego, Calif.).
- Donehower, L. A., et al. (1992). *Nature* 356:215.
- Drummond, I. A., et al. (1994). *Mol. Cell Biol.* 14:3800–9.
- Easton, D., et al. (1993). *Am. J. Hum. Genet.* 52:678–701.
- Eccles, D. M., et al. (1990). *Oncogene* 5:1599–1601.
- Enhancers and Eukaryotic Gene Expression*, Cold Spring Harbor Press, Cold Spring Harbor, N.Y. (1983).
- Erickson, J. et al., (1990). *Science* 249:527–533.
- Fain, P. R. (1992). *Cytogen. Cell Genet.* 60:178.
- Felgner, et al. (1987). *Proc. Natl. Acad. Sci. USA* 84:7413–7417.
- Fiers, et al. (1978). *Nature* 273:113.
- Fink, et al. (1992). *Hum. Gene Ther.* 3:11–19.
- Finkelstein, J., et al. (1990). *Genomics* 7:167–172.
- Freese, et al. (1990). *Biochem. Pharmacol.* 40:2189–2199.
- Friedman, T. (1991). In *Therapy for Genetic Diseases*, T. Friedman, ed., Oxford University Press, pp. 105–121.
- Futreal (1993). Ph.D. Thesis, University of North Carolina, Chapel Hill.
- Futreal, A., et al. (1992a). *Hum. Molec. Genet.* 1:66.
- Futreal, P. A., et al. (1992b). *Cancer Res.* 52:2624–2627.
- Glebov, O. K., et al. (1994). *Cancer Res.* 54:3703–3709.
- Glover, D. (1985). *DNA Cloning*, I and II (Oxford Press).
- Go, R. C. P., et al. (1983). *J. Natl. Cancer Inst.* 71:455–461.
- Goding (1986). *Monoclonal Antibodies: Principles and Practice*, 2d ed. (Academic Press, N.Y.).
- Godowski, et al. (1988). *Science* 241:812–816.
- Goldgar, D. E., et al. (1994). *J. Natl. Can. Inst.* 86:3:200–209.
- Gordon, et al. (1980). *Proc. Natl. Acad. Sci. USA* 77:7380–7384.
- Gorziglia and Kapikian (1992). *J. Virol.* 66:4407–4412.
- Graham and van der Eb (1973). *Virology* 52:456–467.
- Grompe, M., (1993). *Nature Genetics* 5:111–117.
- Grompe, M., et al., (1989). *Proc. Natl. Acad. Sci. USA* 86:5855–5892.
- Guthrie, G. & Fink, G. R. (1991). *Guide to Yeast Genetics and Molecular Biology* (Academic Press).
- Haber, D. A., et al. (1990). *Cell* 61:1257–69.
- Hall, J. M., et al. (1990). *Science* 250:1684–1689.
- Hall, J. M., et al. (1992). *Am. J. Hum. Genet.* 50:1235–1241.
- Harlow & Lane (1988). *Antibodies: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.).
- Hasty, P., K., et al. (1991). *Nature* 350:243.
- Helseth, et al. (1990). *J. Virol.* 64:2416–2420.
- Hodgson, J. (1991). *BioTechnology* 9:19–21.
- Huse, et al. (1989). *Science* 246:1275–1281.
- Innis et al. (1990). *PCR Protocols. A Guide to Methods and Applications* (Academic Press, San Diego, Calif.).
- Jablonski, E., et al. (1986). *Nuc. Acids Res.* 14:6115–6128.
- Jacobs, I. J., et al. (1993). *Cancer Res.* 53:1218–1221.
- Jakoby, W. B. and Pastan, I. H. (eds.) (1979). *Cell Culture. Methods in Enzymology*, volume 58 (Academic Press, Inc., Harcourt Brace Jovanovich (New York)).
- Jeffreys, et al. (1985). *Nature* 314:67–73.
- Johnson, et al. (1992). *J. Virol.* 66:2952–2965.
- Kamb, A. et al. (1994). *Science* 264:436–440.
- Kandpal, et al. (1990). *Nucl. Acids Res.* 18:1789–1795.
- Kaneda, et al. (1989). *J. Biol. Chem.* 264:12126–12129.
- Kanehisa (1984). *Nucl. Acids Res.* 12:203–213.
- Kelsell, D. P., et al. (1993). *Human Mol. Genet.* 2:1823–1828.
- Kinszler, K. W., et al. (1991). *Science* 251:1366–1370.
- Knudson, A. G. (1993). *Nature Genet.* 5:103.
- Kohler, G. and Milstein, C. (1975). *Nature* 256:495–497.
- Kozak, M. (1987). *Nucleic Acids Res.* 15:8125–8148.
- Kraemer, F. B. et al. (1993). *J. Lipid Res.* 34:663–672.
- Kubo, T., et al. (1988). *FEBS Letts.* 241:119.
- Landegren, et al. (1988). *Science* 242:229.
- Lim, et al. (1992). *Circulation* 83:2007–2011.
- Lindsay, S., et al. (1987). *Nature* 327:336–368.
- Litt, et al. (1989). *Am. J. Hum. Genet.* 44:397–401.
- Little, M. H., et al. (1992). *Proc. Natl. Acad. Sci. USA* 89:4791.
- Little, M. H., et al. (1993). *Hum. Mol. Genet.* 2:259.
- Lovett, et al. (1991). *Proc. Natl. Acad. Sci. USA* 88:9628–9632.
- Lynch, H. T., et al. (1990). *Gynecol. Oncol.* 36:48–55.
- Madzak, et al. (1992). *J. Gen. Virol.* 73:1533–1536.
- Malkin, D., et al. (1990). *Science* 250:1233–1238.
- Maniatis, T., et al. (1982). *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.).
- Mann and Baltimore (1985). *J. Virol.* 54:401–407.
- Margaritte, et al. (1992). *Am. J. Hum. Genet.* 50:1231–1234.

A000303

5,753,441

67

Margolskee (1992). *Curr. Top. Microbiol. Immunol.* 158:67-90.

Martin, R., et al. (1990). *BioTechniques* 9:762-768.

Matteucci, M. D. and Caruthers, M. H. (1981). *J Am. Chem. Soc.* 103:3185.

Matthews & Kricka (1988). *Anal. Biochem.* 169:1.

Merrifield (1963). *J. Am. Chem. Soc.* 85:2149-2156.

Mettlin, C., et al. (1990). *American Journal of Epidemiology* 131:973-983.

Metzger, et al. (1988). *Nature* 334:31-36.

Miller (1992). *Curr. Top. Microbiol. Immunol.* 158:1-24.

Miller, et al. (1985). *Mol. Cell. Biol.* 5:431-437.

Miller, et al. (1988). *J. Virol.* 62:4337-4345.

Mittlin (1989). *Clinical Chem.* 35:1819.

Modrich, P. (1991). *Ann. Rev. Genet.* 25:229-253.

Mombaerts, P., et al. (1992). *Cell* 68:869.

Monaco, et al. (1986). *Nature* 323:646.

Moss (1992). *Curr. Top. Microbiol. Immunol.* 158:25-38.

Muzyczka (1992). *Curr. Top. Microbiol. Immunol.* 158:97-123.

Nabel (1992). *Hum. Gene Ther.* 3:399-410.

Nabel, et al. (1990). *Science* 249:1285-1288.

Nakamura, et al. (1987). *Science* 235:1616-1622.

Narod, S. A., et al. (1991). *The Lancet* 338:82-83.

Newman, B., et al. (1988). *Proc. Natl. Acad. Sci. USA* 85:3044-3048.

Newton, C. R., Graham, A., Heptinstall, L. E., Powell, S. J., Summers, C., Kalsheker, N., Smith, J. C., and Markham, A. F. (1989). *Nucl. Acids Res.* 17:2503-2516.

Nguyen, Q., et al. (1992). *BioTechniques* 13:116-123.

Novack, et al. (1986). *Proc. Natl. Acad. Sci. USA* 83:586.

Oh, J. (1985). *Analysis of Human Genetic Linkage*, Johns Hopkins University Press, Baltimore, Md., pp. 1-216.

Ohi, et al. (1990). *Gene* 89:279-282.

Oliphant, A., et al. (1991). *Nucleic Acid Res.* 19:4794.

Oliphant, A., et al. (1991). *Nucleic Acid Res.* 19:4795.

Orita, et al. (1989). *Proc. Natl. Acad. Sci. USA* 86:2776-2770.

Page, et al. (1990). *J. Virol.* 64:5370-5276.

Pellicer, et al. (1980). *Science* 209:1414-1422.

Petropoulos, et al. (1992). *J. Virol.* 66:3391-3397.

Philpott, K. L., et al. (1992). *Science* 256:1448.

Pierce, et al. (1992). *Proc. Natl. Acad. Sci. USA* 89:2056-2060.

Quantin, et al. (1992). *Proc. Natl. Acad. Sci. USA* 89:2581-2584.

Rano & Kidd (1989). *Nucl. Acids Res.* 17:8392.

Rigby, P. W. J., et al. (1977). *J. Mol. Biol.* 113:237-251.

Rosenfeld, et al. (1992). *Cell* 68:143-155.

Sambrook, J., et al. (1989). *Molecular Cloning. A Laboratory Manual*, 2nd Ed. (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.).

Sato, T., et al. (1990). *Cancer Res.* 50:7184-7189.

Scharf (1986). *Science* 233:1076.

Scopes, R. (1982). *Protein Purification: Principles and Practice*, (Springer-Verlag, N.Y.).

Shaulian, E., et al. (1992). *Mol. Cell Biol.* 12:5581-92.

Sheffield, V. C., et al. (1989). *Proc. Natl. Acad. Sci. USA* 86:232-236.

Sheffield, V. C., et al. (1991). *Am. J. Hum. Genet.* 49:699-706.

Shenk, et al. (1975). *Proc. Natl. Acad. Sci. USA* 72:989.

Shimada, et al. (1991). *J. Clin. Invest.* 88:1043-1047.

Shinkai, Y., et al. (1992). *Cell* 68:855.

Shizuya, H., et al. (1992). *Proc. Natl. Acad. Sci. USA* 89:8794-8797.

Simard, J., et al. (1993). *Human Mol. Genet.* 2:1193-1199.

68

Skolnick, M. H. and Wallace, B. R. (1988). *Genomics* 2:273-279.

Skolnick, M. H., et al. (1990). *Science* 250:1715-1720.

Smith, S. A., et al. (1992). *Nature Genetics* 2:128-131.

Smith, T. F. and Waterman, M. S. (1981). *J. Mol. Biol.* 147:195-197.

Snouwvaert, J. N., et al. (1992). *Science* 257:1083.

Sorge, et al. (1984). *Mol. Cell. Biol.* 4:1730-1737.

Srivastava, S., et al. (1993). *Cancer Res.* 53:4452-5.

Sternberg (1990). *Proc. Natl. Acad. Sci. USA* 87:103-107.

Sternberg, et al. (1990). *The New Biologist* 2:151-162.

Stewart, et al. (1992). *Hum. Gene Ther.* 3:267-275.

Stratford-Perricaudet, et al. (1990). *Hum. Gene Ther.* 1:241-256.

Swift, M., et al. (1991). *N. Engl. J. Med.* 325:1831-1836.

Swift, M., et al. (1976). *Cancer Res.* 36:209-215.

Su, L. K., et al. (1993). *Cancer Res.* 53:2728-31.

Thomas, A. and Skolnick, M. H. (1994). *IMA Journal of Mathematics Applied in Medicine and Biology* (in press).

Tonolio, D., et al. (1990). Cold Spring Harbor Conference.

Valancius, V. & Smithies, O. (1991). *Mol. Cell Biol.* 11:1402.

van Dilla, et al. (1986). *Biotechnology* 4:537-552.

Wagner, et al. (1990). *Proc. Natl. Acad. Sci. USA* 87:3410-3414.

Wagner, et al. (1991). *Proc. Natl. Acad. Sci. USA* 88:4255-4259.

Wang and Huang (1989). *Biochemistry* 28:9508-9514.

Wartell, R. M., et al. (1990). *Nucl. Acids Res.* 18:2699-2705.

Weber, J. L. (1990). *Genomics* 7:524-530.

Weber and May (1989). *Am. J. Hum. Genet.* 44:388-396.

Weber, J. L., et al. (1990). *Nucleic Acid Res.* 18:4640.

Wells, J. A. (1991). *Methods in Enzymol* 202:390-411.

Wetmur & Davidson (1968). *J. Mol. Biol.* 31:349-370.

White, M. B., et al., (1992). *Genomics* 12:301-306.

White and Lalouel (1988). *Ann. Rev. Genet.* 22:259-279.

Wilkinson, et al. (1992). *Nucleic Acids Res.* 20:2233-2239.

Williams and Anderson (1984). *Genet. Epidemiol.* 1:7-20.

Wolff, et al. (1990). *Science* 247:1465-1468.

Wolff, et al. (1991). *BioTechniques* 11:474-485.

Wooster, R., et al. (1994). *Science* 265:2088.

Wu, et al. (1989a). *Genomics* 4:560-569.

Wu, et al. (1989b). *J. Biol. Chem.* 264:16985-16987.

Wu, et al. (1991). *J. Biol. Chem.* 266:14338-14342.

Zenke, et al. (1990). *Proc. Natl. Acad. Sci. USA* 87:3655-3659.

List of Patents and Patent Applications:

U.S. Pat. No. 3,817,837

U.S. Pat. No. 3,850,752

U.S. Pat. No. 3,939,350

U.S. Pat. No. 3,996,345

U.S. Pat. No. 4,275,149

U.S. Pat. No. 4,277,437

U.S. Pat. No. 4,366,241

U.S. Pat. No. 4,376,110

U.S. Pat. No. 4,486,530

U.S. Pat. No. 4,683,195

U.S. Pat. No. 4,683,202

U.S. Pat. No. 4,816,567

U.S. Pat. No. 4,868,105

U.S. Pat. No. 5,252,479

EPO Publication No. 225,807

European Patent Application Publication No. 0332435

Geysen, H., PCT published application WO 84/03564, published 13 Sep. 1984

Hitzeman et al., EP 73,675A

PCT published application WO 93/07282

A000304

5,753,441

69

70

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i i i) NUMBER OF SEQUENCES: 85

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5914 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: cDNA

(i i i) HYPOTHETICAL: NO

(i v) ANTI-SENSE: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(i x) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 120..5708

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```
AGCTCGCTGA GACTTCCTGG ACCCCGCACC AGGCTGTGGG GTTTCCTAGA TAAC TGGGCC 60
CCTGCGCTCA GGAGGCCCTC ACCCTCTGCT CTGGGTAAAG TTCATTGGAA CAGAAAGAA 119
ATG GAT TTA TCT GCT CTT CGC GTT GAA GAA GTA CAA AAT GTC ATT AAT 167
Met Asp Leu Ser Ala Leu Arg Val Glu Glu Val Gln Asn Val Ile Asn
1 5 10 15
GCT ATG CAG AAA ATC TTA GAG TGT CCC ATC TGT CTG GAG TTG ATC AAG 215
Ala Met Gln Lys Ile Leu Glu Cys Pro Ile Cys Leu Glu Leu Ile Lys
20 25 30
GAA CCT GTC TCC ACA AAG TGT GAC CAC ATA TTT TGC AAA TTT TGC ATG 263
Glu Pro Val Ser Thr Lys Cys Asp His Ile Phe Cys Lys Phe Cys Met
35 40 45
CTG AAA CTT CTC AAC CAG AAG AAA GGG CCT TCA CAG TGT CCT TTA TGT 311
Leu Lys Leu Leu Asn Gln Lys Lys Gly Pro Ser Gln Cys Pro Leu Cys
50 55 60
AAG AAT GAT ATA ACC AAA AGG AGC CTA CAA GAA AGT ACG AGA TTT AGT 359
Lys Asn Asp Ile Thr Lys Arg Ser Leu Gln Glu Ser Thr Arg Phe Ser
65 70 75 80
CAA CTT GTT GAA GAG CTA TTG AAA ATC ATT TGT GCT TTT CAG CTT GAC 407
Gln Leu Val Glu Glu Leu Leu Lys Ile Ile Cys Ala Phe Gln Leu Asp
85 90 95
ACA GGT TTG GAG TAT GCA AAC AGC TAT AAT TTT GCA AAA AAG GAA AAT 455
Thr Gly Leu Glu Tyr Ala Asn Ser Tyr Asn Phe Ala Lys Lys Glu Asn
100 105 110
AAC TCT CCT GAA CAT CTA AAA GAT GAA GTT TCT ATC ATC CAA AGT ATG 503
Asn Ser Pro Glu His Leu Lys Asp Glu Val Ser Ile Ile Gln Ser Met
115 120 125
GGC TAC AGA AAC CGT GCC AAA AGA CTT CTA CAG AGT GAA CCC GAA AAT 551
Gly Tyr Arg Asn Arg Ala Lys Arg Leu Leu Gln Ser Glu Pro Glu Asn
130 135 140
CCT TCC TTG CAG GAA ACC AGT CTC AGT GTC CAA CTC TCT AAC CTT GGA 599
Pro Ser Leu Gln Glu Thr Ser Leu Ser Val Gln Leu Ser Asn Leu Gly
145 150 155 160
ACT GTG AGA ACT CTG AGG ACA AAG CAG CGG ATA CAA CCT CAA AAG ACG 647
Thr Val Arg Thr Leu Arg Thr Lys Gln Arg Ile Gln Pro Gln Lys Thr
165 170 175
TCT GTC TAC ATT GAA TTG GGA TCT GAT TCT TCT GAA GAT ACC GTT AAT 695
```

5,753,441

71

72

-continued

Ser	Val	Tyr	Ile	Glu	Leu	Gly	Ser	Asp	Ser	Ser	Glu	Asp	Thr	Val	Asn	
			180					185					190			
AAG	GCA	ACT	TAT	TGC	AGT	GTG	GGA	GAT	CAA	GAA	TTG	TTA	CAA	ATC	ACC	743
Lys	Ala	Thr	Tyr	Cys	Ser	Val	Gly	Asp	Gln	Glu	Leu	Leu	Gln	Ile	Thr	
		195					200					205				
CCT	CAA	GGA	ACC	AGG	GAT	GAA	ATC	AGT	TTG	GAT	TCT	GCA	AAA	AAG	GCT	791
Pro	Gln	Gly	Thr	Arg	Asp	Glu	Ile	Ser	Leu	Asp	Ser	Ala	Lys	Lys	Ala	
	210					215					220					
GCT	TGT	GAA	TTT	TCT	GAG	ACG	GAT	GTA	ACA	AAT	ACT	GAA	CAT	CAT	CAA	839
Ala	Cys	Glu	Phe	Ser	Glu	Thr	Asp	Val	Thr	Asn	Thr	Glu	His	His	Gln	
	225				230					235					240	
CCC	AGT	AAT	AAT	GAT	TTG	AAC	ACC	ACT	GAG	AAG	CGT	GCA	GCT	GAG	AGG	887
Pro	Ser	Asn	Asn	Asp	Leu	Asn	Thr	Thr	Glu	Lys	Arg	Ala	Ala	Glu	Arg	
				245					250					255		
CAT	CCA	GAA	AAG	TAT	CAG	GGT	AGT	TCT	GIT	TCA	AAC	TTG	CAT	GTG	GAG	935
His	Pro	Glu	Lys	Tyr	Gln	Gly	Ser	Ser	Val	Ser	Asn	Leu	His	Val	Glu	
			260					265					270			
CCA	TGT	GGC	ACA	AAT	ACT	CAT	GCC	AGC	TCA	TTA	CAG	CAT	GAG	AAC	AGC	983
Pro	Cys	Gly	Thr	Asn	Thr	His	Ala	Ser	Ser	Leu	Gln	His	Glu	Asn	Ser	
		275					280					285				
AGT	TTA	TTA	CTC	ACT	AAA	GAC	AGA	ATG	AAT	GTA	GAA	AAG	GCT	GAA	TTC	1031
Ser	Leu	Leu	Leu	Thr	Lys	Asp	Arg	Met	Asn	Val	Glu	Lys	Ala	Glu	Phe	
	290					295					300					
TGT	AAT	AAA	AGC	AAA	CAG	CCT	GGC	TTA	GCA	AGG	AGC	CAA	CAT	AAC	AGA	1079
Cys	Asn	Lys	Ser	Lys	Gln	Pro	Gly	Leu	Ala	Arg	Ser	Gln	His	Asn	Arg	
	305				310					315					320	
TGG	GCT	GGA	AGT	AAG	GAA	ACA	TGT	AAT	GAT	AGG	CGG	ACT	CCC	AGC	ACA	1127
Trp	Ala	Gly	Ser	Lys	Glu	Thr	Cys	Asn	Asp	Arg	Arg	Thr	Pro	Ser	Thr	
				325					330					335		
GAA	AAA	AAG	GTA	GAT	CTG	AAT	GCT	GAT	CCC	CTG	TGT	GAG	AGA	AAA	GAA	1175
Glu	Lys	Lys	Val	Asp	Leu	Asn	Ala	Asp	Pro	Leu	Cys	Glu	Arg	Lys	Glu	
			340					345					350			
TGG	AAT	AAG	CAG	AAA	CTG	CCA	TGC	TCA	GAG	AAT	CCT	AGA	GAT	ACT	GAA	1223
Trp	Asn	Lys	Gln	Lys	Leu	Pro	Cys	Ser	Glu	Asn	Pro	Arg	Asp	Thr	Glu	
		355					360					365				
GAT	GTT	CCT	TGG	ATA	ACA	CTA	AAT	AGC	AGC	ATT	CAG	AAA	GTT	AAT	GAG	1271
Asp	Val	Pro	Trp	Ile	Thr	Leu	Asn	Ser	Ser	Ile	Gln	Lys	Val	Asn	Glu	
	370					375					380					
TGG	TTT	TCC	AGA	AGT	GAT	GAA	CTG	TTA	GGT	TCT	GAT	GAC	TCA	CAT	GAT	1319
Trp	Phe	Ser	Arg	Ser	Asp	Glu	Leu	Leu	Gly	Ser	Asp	Asp	Ser	His	Asp	
	385				390					395					400	
GGG	GAG	TCT	GAA	TCA	AAT	GCC	AAA	GTA	GCT	GAT	GTA	TTG	GAC	GTT	CTA	1367
Gly	Glu	Ser	Glu	Ser	Asn	Ala	Lys	Val	Ala	Asp	Val	Leu	Asp	Val	Leu	
				405					410					415		
AAT	GAG	GTA	GAT	GAA	TAT	TCT	GGT	TCT	TCA	GAG	AAA	ATA	GAC	TTA	CTG	1415
Asn	Glu	Val	Asp	Glu	Tyr	Ser	Gly	Ser	Ser	Glu	Lys	Ile	Asp	Leu	Leu	
			420					425					430			
GCC	AGT	GAT	CCT	CAT	GAG	GCT	TTA	ATA	TGT	AAA	AGT	GAA	AGA	GTT	CAC	1463
Ala	Ser	Asp	Pro	His	Glu	Ala	Leu	Ile	Cys	Lys	Ser	Glu	Arg	Val	His	
		435					440					445				
TCC	AAA	TCA	GTA	GAG	AGT	AAT	ATT	GAA	GAC	AAA	ATA	TTT	GGG	AAA	ACC	1511
Ser	Lys	Ser	Val	Glu	Ser	Asn	Ile	Glu	Asp	Lys	Ile	Phe	Gly	Lys	Thr	
	450					455					460					
TAT	CGG	AAG	AAG	GCA	AGC	CTC	CCC	AAC	TTA	AGC	CAT	GTA	ACT	GAA	AAT	1559
Tyr	Arg	Lys	Lys	Ala	Ser	Leu	Pro	Asn	Leu	Ser	His	Val	Thr	Glu	Asn	
	465				470					475					480	
CTA	ATT	ATA	GGA	GCA	TTT	GTT	ACT	GAG	CCA	CAG	ATA	ATA	CAA	GAG	CGT	1607
Leu	Ile	Ile	Gly	Ala	Phe	Val	Thr	Glu	Pro	Gln	Ile	Ile	Gln	Glu	Arg	
				485					490					495		
CCC	CTC	ACA	AAT	AAA	TTA	AAG	CGT	AAA	AGG	AGA	CCT	ACA	TCA	GGC	CTT	1655

A000306

5,753,441

73

74

-continued

Pro	Leu	Thr	Asn	Lys	Leu	Lys	Arg	Lys	Arg	Arg	Pro	Thr	Ser	Gly	Leu	
			500					505					510			
CAT	CCT	GAG	GAT	TTT	ATC	AAG	AAA	GCA	GAT	TTG	GCA	GTT	CAA	AAG	ACT	1703
His	Pro	Glu	Asp	Phe	Ile	Lys	Lys	Ala	Asp	Leu	Ala	Val	Gln	Lys	Thr	
		515					520					525				
CCT	GAA	ATG	ATA	AAT	CAG	GGA	ACT	AAC	CAA	ACG	GAG	CAG	AAT	GGT	CAA	1751
Pro	Glu	Met	Ile	Asn	Gln	Gly	Thr	Asn	Gln	Thr	Glu	Gln	Asn	Gly	Gln	
		530				535					540					
GTG	ATG	AAT	ATT	ACT	AAT	AGT	GGT	CAT	GAG	AAT	AAA	ACA	AAA	GGT	GAT	1799
Val	Met	Asn	Ile	Thr	Asn	Ser	Gly	His	Glu	Asn	Lys	Thr	Lys	Gly	Asp	
					550				555						560	
TCT	ATT	CAG	AAT	GAG	AAA	AAT	CCT	AAC	CCA	ATA	GAA	TCA	CTC	GAA	AAA	1847
Ser	Ile	Gln	Asn	Glu	Lys	Asn	Pro	Asn	Pro	Ile	Glu	Ser	Leu	Glu	Lys	
				565					570					575		
GAA	TCT	GCT	TTC	AAA	ACG	AAA	GCT	GAA	CCT	ATA	AGC	AGC	AGT	ATA	AGC	1895
Glu	Ser	Ala	Phe	Lys	Thr	Lys	Ala	Glu	Pro	Ile	Ser	Ser	Ser	Ile	Ser	
			580					585					590			
AAT	ATG	GAA	CTC	GAA	TTA	AAT	ATC	CAC	AAT	TCA	AAA	GCA	CCT	AAA	AAG	1943
Asn	Met	Glu	Leu	Glu	Leu	Asn	Ile	His	Asn	Ser	Lys	Ala	Pro	Lys	Lys	
		595					600					605				
AAT	AGG	CTG	AGG	AGG	AAG	TCT	TCT	ACC	AGG	CAT	ATT	CAT	GCG	CTT	GAA	1991
Asn	Arg	Leu	Arg	Arg	Lys	Ser	Ser	Thr	Arg	His	Ile	His	Ala	Leu	Glu	
		610				615					620					
CTA	GTA	GTC	AGT	AGA	AAT	CTA	AGC	CCA	CCT	AAT	TGT	ACT	GAA	TTG	CAA	2039
Leu	Val	Val	Ser	Arg	Asn	Leu	Ser	Pro	Pro	Asn	Cys	Thr	Glu	Leu	Gln	
					630					635					640	
ATT	GAT	AGT	TGT	TCT	AGC	AGT	GAA	GAG	ATA	AAG	AAA	AAA	AAG	TAC	AAC	2087
Ile	Asp	Ser	Cys	Ser	Ser	Ser	Glu	Glu	Ile	Lys	Lys	Lys	Lys	Tyr	Asn	
			645					650					655			
CAA	ATG	CCA	GTC	AGG	CAC	AGC	AGA	AAC	CTA	CAA	CTC	ATG	GAA	GGT	AAA	2135
Gln	Met	Pro	Val	Arg	His	Ser	Arg	Asn	Leu	Gln	Leu	Met	Glu	Gly	Lys	
			660					665					670			
GAA	CCT	GCA	ACT	GGA	GCC	AAG	AAG	AGT	AAC	AAG	CCA	AAT	GAA	CAG	ACA	2183
Glu	Pro	Ala	Thr	Gly	Ala	Lys	Lys	Ser	Asn	Lys	Pro	Asn	Glu	Gln	Thr	
		675					680					685				
AGT	AAA	AGA	CAT	GAC	AGC	GAT	ACT	TTC	CCA	GAG	CTG	AAG	TTA	ACA	AAT	2231
Ser	Lys	Arg	His	Asp	Ser	Asp	Thr	Phe	Pro	Glu	Leu	Lys	Leu	Thr	Asn	
		690				695					700					
GCA	CCT	GGT	TCT	TTT	ACT	AAG	TGT	TCA	AAT	ACC	AGT	GAA	CTT	AAA	GAA	2279
Ala	Pro	Gly	Ser	Phe	Thr	Lys	Cys	Ser	Asn	Thr	Ser	Glu	Leu	Lys	Glu	
		705			710					715					720	
TTT	GTC	AAT	CCT	AGC	CTT	CCA	AGA	GAA	GAA	AAA	GAA	GAG	AAA	CTA	GAA	2327
Phe	Val	Asn	Pro	Ser	Leu	Pro	Arg	Glu	Glu	Lys	Glu	Glu	Lys	Leu	Glu	
			725					730					735			
ACA	GTT	AAA	GTC	TCT	AAT	AAT	GCT	GAA	GAC	CCC	AAA	GAT	CTC	ATG	TTA	2375
Thr	Val	Lys	Val	Ser	Asn	Asn	Ala	Glu	Asp	Pro	Lys	Asp	Leu	Met	Leu	
			740					745					750			
AGT	GGA	GAA	AGG	GTT	TTG	CAA	ACT	GAA	AGA	TCT	GTA	GAG	AGT	AGC	AGT	2423
Ser	Gly	Glu	Arg	Val	Leu	Gln	Thr	Glu	Arg	Ser	Val	Glu	Ser	Ser	Ser	
		755					760					765				
ATT	TCA	TTG	GTA	CCT	GGT	ACT	GAT	TAT	GGC	ACT	CAG	GAA	AGT	ATC	TCG	2471
Ile	Ser	Leu	Val	Pro	Gly	Thr	Asp	Tyr	Gly	Thr	Gln	Glu	Ser	Ile	Ser	
		770			775						780					
TTA	CTG	GAA	GTT	AGC	ACT	CTA	GGG	AAG	GCA	AAA	ACA	GAA	CCA	AAT	AAA	2519
Leu	Leu	Glu	Val	Ser	Thr	Leu	Gly	Lys	Ala	Lys	Thr	Glu	Pro	Asn	Lys	
		785			790					795					800	
TGT	GTG	AGT	CAG	TGT	GCA	GCA	TTT	GAA	AAC	CCC	AAG	GGA	CTA	ATT	CAT	2567
Cys	Val	Ser	Gln	Cys	Ala	Ala	Phe	Glu	Asn	Pro	Lys	Gly	Leu	Ile	His	
			805					810					815			
GGT	TGT	TCC	AAA	GAT	AAT	AGA	AAT	GAC	ACA	GAA	GGC	TIT	AAG	TAT	CCA	2615

A000307

5,753,441

75

76

-continued

Gly	Cys	Ser	Lys	Asp	Asn	Arg	Asn	Asp	Thr	Glu	Gly	Phe	Lys	Tyr	Pro	
			820					825					830			
TTG	GGA	CAT	GAA	GTT	AAC	CAC	AGT	CGG	GAA	ACA	AGC	ATA	GAA	ATG	GAA	2663
Leu	Gly	His	Glu	Val	Asn	His	Ser	Arg	Glu	Thr	Ser	Ile	Glu	Met	Glu	
		835					840					845				
GAA	AGT	GAA	CTT	GAT	GCT	CAG	TAT	TTG	CAG	AAT	ACA	TTC	AAG	GTT	TCA	2711
Glu	Ser	Glu	Leu	Asp	Ala	Gln	Tyr	Leu	Gln	Asn	Thr	Phe	Lys	Val	Ser	
	850					855					860					
AAG	CGC	CAG	TCA	TTT	GCT	CCG	TTT	TCA	AAT	CCA	GGA	AAT	GCA	GAA	GAG	2759
Lys	Arg	Gln	Ser	Phe	Ala	Pro	Phe	Ser	Asn	Pro	Gly	Asn	Ala	Glu	Glu	
	865				870					875					880	
GAA	TGT	GCA	ACA	TTC	TCT	GCC	CAC	TCT	GGG	TCC	TTA	AAG	AAA	CAA	AGT	2807
Glu	Cys	Ala	Thr		Ser	Ala	His	Ser	Gly	Ser	Leu	Lys	Lys	Gln	Ser	
				885					890					895		
CCA	AAA	GTC	ACT	TTT	GAA	TGT	GAA	CAA	AAG	GAA	GAA	AAT	CAA	GGA	AAG	2855
Pro	Lys	Val	Thr	Phe	Glu	Cys	Glu	Gln	Lys	Glu	Glu	Asn	Gln	Gly	Lys	
		900						905					910			
AAT	GAG	TCT	AAT	ATC	AAG	CCT	GTA	CAG	ACA	GTT	AAT	ATC	ACT	GCA	GGC	2903
Asn	Glu	Ser	Asn	Ile	Lys	Pro	Val	Gln	Thr	Val	Asn	Ile	Thr	Ala	Gly	
		915					920					925				
TTT	CCT	GTG	GTT	GGT	CAG	AAA	GAT	AAG	CCA	GTT	GAT	AAT	GCC	AAA	TGT	2951
Phe	Pro	Val	Val	Gly	Gln	Lys	Asp	Lys	Pro	Val	Asp	Asn	Ala	Lys	Cys	
	930					935					940					
AGT	ATC	AAA	GGA	GGC	TCT	AGG	TTT	TGT	CTA	TCA	TCT	CAG	TTC	AGA	GGC	2999
Ser	Ile	Lys	Gly	Gly		Arg	Phe	Cys	Leu		Ser	Gln	Phe	Arg	Gly	
	945				950					955					960	
AAC	GAA	ACT	GGA	CTC	ATT	ACT	CCA	AAT	AAA	CAT	GGA	CTT	TTA	CAA	AAC	3047
Asn	Glu	Thr	Gly	Leu	Ile	Thr	Pro	Asn	Lys	His	Gly	Leu	Leu	Gln	Asn	
				965					970					975		
CCA	TAT	CGT	ATA	CCA	CCA	CTT	TTT	CCC	ATC	AAG	TCA	TTT	GTT	AAA	ACT	3095
Pro	Tyr	Arg	Ile	Pro	Pro	Leu	Phe	Pro	Ile	Lys	Ser	Phe	Val	Lys	Thr	
			980					985					990			
AAA	TGT	AAG	AAA	AAT	CTG	CTA	GAG	GAA	AAC	TTT	GAG	GAA	CAT	TCA	ATG	3143
Lys	Cys	Lys	Lys	Asn	Leu	Leu	Glu	Glu	Asn	Phe	Glu	Glu	His	Ser	Met	
		995					1000					1005				
TCA	CCT	GAA	AGA	GAA	ATG	GGA	AAT	GAG	AAC	ATT	CCA	AGT	ACA	GTG	AGC	3191
Ser	Pro	Glu	Arg	Glu	Met	Gly	Asn	Glu	Asn	Ile	Pro	Ser	Thr	Val	Ser	
	1010					1015					1020					
ACA	ATT	AGC	CGT	AAT	AAC	ATT	AGA	GAA	AAT	GTT	TTT	AAA	GAA	GCC	AGC	3239
Thr	Ile	Ser	Arg	Asn	Asn	Ile	Arg	Glu	Asn	Val	Phe	Lys	Glu	Ala	Ser	
	1025				1030					1035					1040	
TCA	AGC	AAT	ATT	AAT	GAA	GTA	GGT	TCC	AGT	ACT	AAT	GAA	GTG	GGC	TCC	3287
Ser	Ser	Asn	Ile	Asn	Glu	Val	Gly	Ser	Ser	Thr	Asn	Glu	Val	Gly	Ser	
				1045					1050					1055		
AGT	ATT	AAT	GAA	ATA	GGT	TCC	AGT	GAT	GAA	AAC	ATT	CAA	GCA	GAA	CTA	3335
Ser	Ile	Asn	Glu	Ile	Gly	Ser	Ser	Asp	Glu	Asn	Ile	Gln	Ala	Glu	Leu	
			1060					1065					1070			
GGT	AGA	AAC	AGA	GGG	CCA	AAA	TTG	AAT	GCT	ATG	CTT	AGA	TTA	GGG	GTT	3383
Gly	Arg	Asn	Arg	Gly	Pro	Lys	Leu	Asn	Ala	Met	Leu	Arg	Leu	Gly	Val	
		1075					1080					1085				
TTG	CAA	CCT	GAG	GTC	TAT	AAA	CAA	AGT	CIT	CCT	GGA	AGT	AAT	TGT	AAG	3431
Leu	Gln	Pro	Glu	Val	Tyr	Lys	Gln	Ser	Leu	Pro	Gly	Ser	Asn	Cys	Lys	
	1090					1095					1100					
CAT	CCT	GAA	ATA	AAA	AAG	CAA	GAA	TAT	GAA	GAA	GTA	GTT	CAG	ACT	GTT	3479
His	Pro	Glu	Ile	Lys	Lys	Gln	Glu	Tyr	Glu	Glu	Val	Val	Gln	Thr	Val	
	1105				1110					1115					1120	
AAT	ACA	GAT	TTC	TCT	CCA	TAT	CTG	ATT	TCA	GAT	AAC	TTA	GAA	CAG	CCT	3527
Asn	Thr	Asp	Phe	Ser	Pro	Tyr	Leu	Ile	Ser	Asp	Asn	Leu	Glu	Gln	Pro	
				1125					1130					1135		
ATG	GGA	AGT	AGT	CAT	GCA	TCT	CAG	GTT	TGT	TCT	GAG	ACA	CCT	GAT	GAC	3575

A000308

5,753,441

77

78

-continued

Met	Gly	Ser	Ser	His	Ala	Ser	Gln	Val	Cys	Ser	Glu	Thr	Pro	Asp	Asp	
			1140					1145					1150			
CTG Leu	TTA Leu	GAT Asp	GAT Asp	GGT Gly	GAA Glu	ATA Ile	AAG Lys	GAA Glu	GAT Asp	ACT Thr	AGT Ser	TTT Phe	GCT Ala	GAA Glu	AAT Asn	3623
GAC Asp	ATT Ile	AAG Lys	GAA Glu	AGT Ser	TCT Ser	GCT Ala	GTT Val	TTT Phe	AGC Ser	AAA Lys	AGC Ser	GTC Val	CAG Gln	AAA Lys	GGA Gly	3671
GAG Glu	CTT Leu	AGC Ser	AGG Arg	AGT Ser	CCT Pro	AGC Ser	CCT Pro	TTC Phe	ACC Thr	CAT His	ACA Thr	CAT His	TTG Leu	GCT Ala	CAG Gln	3719
GGT Gly	TAC Tyr	CGA Arg	AGA Arg	GGG Gly	GCC Ala	AAG Lys	AAA Lys	TTA Leu	GAG Glu	TCC Ser	TCA Ser	GAA Glu	GAG Glu	AAC Asn	TTA Leu	3767
TCT Ser	AGT Ser	GAG Glu	GAT Asp	GAA Glu	GAG Glu	CTT Leu	CCC Pro	TGC Cys	TTC Phe	CAA Gln	CAC His	TTG Leu	TTA Leu	TTT Phe	GGT Gly	3815
AAA Lys	GTA Val	AAC Asn	AAT Asn	ATA Ile	CCT Pro	TCT Ser	CAG Gln	TCT Ser	ACT Thr	AGG Arg	CAT His	AGC Ser	ACC Thr	GTT Val	GCT Ala	3863
ACC Thr	GAG Glu	TGT Cys	CTG Leu	TCT Ser	AAG Lys	AAC Asn	ACA Thr	GAG Glu	GAG Glu	AAT Asn	TTA Leu	TTA Leu	TCA Ser	TTG Leu	AAG Lys	3911
AAT Asn	AGC Ser	TTA Leu	AAT Asn	GAC Asp	TGC Cys	AGT Ser	AAC Asn	CAG Gln	GTA Val	ATA Ile	TTG Leu	GCA Ala	AAG Lys	GCA Ala	TCT Ser	3959
CAG Gln	GAA Glu	CAT His	CAC His	CTT Leu	AGT Ser	GAG Glu	GAA Glu	ACA Thr	AAA Cys	TGT Cys	TCT Ser	GCT Ala	AGC Ser	TTG Leu	TTT Phe	4007
TCT Ser	TCA Ser	CAG Gln	TGC Cys	AGT Ser	GAA Glu	TTG Leu	GAA Glu	GAC Asp	TTG Leu	ACT Thr	GCA Ala	AAT Asn	ACA Thr	AAC Asn	ACC Thr	4055
CAG Gln	GAT Asp	CCT Pro	TTC Phe	TTG Leu	ATT Ile	GGT Gly	TCT Ser	TCC Ser	AAA Lys	CAA Gln	ATG Met	AGG Arg	CAT His	CAG Gln	TCT Ser	4103
GAA Glu	AGC Ser	CAG Gln	GGA Gly	GTT Val	GGT Gly	CTG Leu	AGT Ser	GAC Asp	AAG Lys	GAA Glu	TTG Leu	GTT Val	TCA Ser	GAT Asp	GAT Asp	4151
GAA Glu	GAA Glu	AGA Arg	GGA Gly	ACG Thr	GGC Gly	TTG Leu	GAA Glu	GAA Glu	AAT Asn	AAT Asn	CAA Gln	GAA Glu	GAG Glu	CAA Gln	AGC Ser	4199
ATG Met	GAT Asp	TCA Ser	AAC Asn	TTA Leu	GGT Gly	GAA Glu	GCA Ala	GCA Ala	TCT Ser	GGG Gly	TGT Cys	GAG Glu	AGT Ser	GAA Glu	ACA Thr	4247
AGC Ser	GTC Val	TCT Ser	GAA Glu	GAC Asp	TGC Cys	TCA Ser	GGG Gly	CTA Leu	TCC Ser	TCT Ser	CAG Gln	AGT Ser	GAC Asp	ATT Ile	TTA Leu	4295
ACC Thr	ACT Thr	CAG Gln	CAG Gln	AGG Arg	GAT Asp	ACC Thr	ATG Met	CAA Gln	CAT His	AAC Asn	CTG Leu	ATA Ile	AAG Lys	CTC Leu	CAG Gln	4343
CAG Gln	GAA Glu	ATG Met	GCT Ala	GAA Glu	CTA Leu	GAA Glu	GCT Ala	GTG Val	TTA Leu	GAA Glu	CAG Gln	CAT His	GGG Gly	AGC Ser	CAG Gln	4391
CCT Pro	TCT Ser	AAC Asn	AGC Ser	TAC Tyr	CCT Pro	TCC Ser	ATC Ile	ATA Ile	AGT Ser	GAC Asp	TCT Ser	TCT Ser	GCC Ala	CTT Leu	GAG Glu	4439
GAC Asp	CTG Leu	CGA Arg	AAT Asn	CCA Pro	GAA Glu	CAA Gln	AGC Ser	ACA Thr	TCA Ser	GAA Glu	AAA Lys	GCA Ala	GTA Val	TTA Leu	ACT Thr	4487
TCA	CAG	AAA	AGT	AGT	GAA	TAC	CCT	ATA	AGC	CAG	AAT	CCA	GAA	GGC	CTT	4535

A000309

5,753,441

79

80

-continued

Ser	Gln	Lys	Ser	Ser	Glu	Tyr	Pro	Ile	Ser	Gln	Asn	Pro	Glu	Gly	Leu	
			1460					1465					1470			
TCT	GCT	GAC	AAG	TTT	GAG	GTG	TCT	GCA	GAT	AGT	TCT	ACC	AGT	AAA	AAT	4583
Ser	Ala	Asp	Lys	Phe	Glu	Val	Ser	Ala	Asp	Ser	Ser	Thr	Ser	Lys	Asn	
		1475					1480					1485				
AAA	GAA	CCA	GGA	GTG	GAA	AGG	TCA	TCC	CCT	TCT	AAA	TGC	CCA	TCA	TTA	4631
Lys	Glu	Pro	Gly	Val	Glu	Arg	Ser	Ser	Pro	Ser	Lys	Cys	Pro	Ser	Leu	
	1490					1495					1500					
GAT	GAT	AGG	TGG	TAC	ATG	CAC	AGT	TGC	TCT	GGG	AGT	CTT	CAG	AAT	AGA	4679
Asp	Asp	Arg	Trp	Tyr	Met	His	Ser	Cys	Ser	Gly	Ser	Leu	Gln	Asn	Arg	
1505					1510					1515					1520	
AAC	TAC	CCA	TCT	CAA	GAG	GAG	CTC	ATT	AAG	GTT	GTT	GAT	GTG	GAG	GAG	4727
Asn	Tyr	Pro	Ser	Gln	Glu	Glu	Leu	Ile	Lys	Val	Val	Asp	Val	Glu	Glu	
				1525					1530					1535		
CAA	CAG	CTG	GAA	GAG	TCT	GGG	CCA	CAC	GAT	TTG	ACG	GAA	ACA	TCT	TAC	4775
Gln	Gln	Leu	Glu	Glu	Ser	Gly	Pro	His	Asp	Leu	Thr	Glu	Thr	Ser	Tyr	
		1540					1545						1550			
TTG	CCA	AGG	CAA	GAT	CTA	GAG	GGA	ACC	CCT	TAC	CTG	GAA	TCT	GGA	ATC	4823
Leu	Pro	Arg	Gln	Asp	Leu	Glu	Gly	Thr	Pro	Tyr	Leu	Glu	Ser	Gly	Ile	
		1555					1560					1565				
AGC	CTC	TTC	TCT	GAT	GAC	CCT	GAA	TCT	GAT	CCT	TCT	GAA	GAC	AGA	GCC	4871
Ser	Leu	Phe	Ser	Asp	Asp	Pro	Glu	Ser	Asp	Pro	Ser	Glu	Asp	Arg	Ala	
		1570				1575					1580					
CCA	GAG	TCA	GCT	CGT	GTT	GGC	AAC	ATA	CCA	TCT	TCA	ACC	TCT	GCA	TTG	4919
Pro	Glu	Ser	Ala	Arg	Val	Gly	Asn	Ile	Pro	Ser	Ser	Thr	Ser	Ala	Leu	
	1585				1590					1595					1600	
AAA	GTT	CCC	CAA	TTG	AAA	GTT	GCA	GAA	TCT	GCC	CAG	AGT	CCA	GCT	GCT	4967
Lys	Val	Pro	Gln	Leu	Lys	Val	Ala	Glu	Ser	Ala	Gln	Ser	Pro	Ala	Ala	
				1605				1610						1615		
GCT	CAT	ACT	ACT	GAT	ACT	GCT	GGG	TAT	AAT	GCA	ATG	GAA	GAA	AGT	GTG	5015
Ala	His	Thr	Thr	Asp	Thr	Ala	Gly	Tyr	Asn	Ala	Met	Glu	Glu	Ser	Val	
		1620						1625					1630			
AGC	AGG	GAG	AAG	CCA	GAA	TTG	ACA	GCT	TCA	ACA	GAA	AGG	GTC	AAC	AAA	5063
Ser	Arg	Glu	Lys	Pro	Glu	Leu	Thr	Ala	Ser	Thr	Glu	Arg	Val	Asn	Lys	
		1635					1640					1645				
AGA	ATG	TCC	ATG	GTG	GTG	TCT	GGC	CTG	ACC	CCA	GAA	GAA	TTT	ATG	CTC	5111
Arg	Met	Ser	Met	Val	Val	Ser	Gly	Leu	Thr	Pro	Glu	Glu	Phe	Met	Leu	
	1650					1655					1660					
GTG	TAC	AAG	TTT	GCC	AGA	AAA	CAC	CAC	ATC	ACT	TTA	ACT	AAT	CTA	ATT	5159
Val	Tyr	Lys	Phe	Ala	Arg	Lys	His	His	Ile	Thr	Leu	Thr	Asn	Leu	Ile	
	1665				1670					1675					1680	
ACT	GAA	GAG	ACT	ACT	CAT	GTT	GTT	ATG	AAA	ACA	GAT	GCT	GAG	TTT	GTG	5207
Thr	Glu	Glu	Thr	Thr	His	Val	Val	Met	Lys	Thr	Asp	Ala	Glu	Phe	Val	
				1685				1690						1695		
TGT	GAA	CGG	ACA	CTG	AAA	TAT	TTT	CTA	GGA	ATT	GCG	GGA	GGA	AAA	TGG	5255
Cys	Glu	Arg	Thr	Leu	Lys	Tyr	Phe	Leu	Gly	Ile	Ala	Gly	Gly	Lys	Trp	
		1700						1705					1710			
GTA	GTT	AGC	TAT	TTC	TGG	GTG	ACC	CAG	ICT	ATT	AAA	GAA	AGA	AAA	ATG	5303
Val	Val	Ser	Tyr	Phe	Trp	Val	Thr	Gln	Ser	Ile	Lys	Glu	Arg	Lys	Met	
		1715					1720					1725				
CTG	AAT	GAG	CAT	GAT	TTT	GAA	GTC	AGA	GGA	GAT	GTG	GTC	AAT	GGA	AGA	5351
Leu	Asn	Glu	His	Asp	Phe	Glu	Val	Arg	Gly	Asp	Val	Val	Asn	Gly	Arg	
	1730					1735					1740					
AAC	CAC	CAA	GGT	CCA	AAG	CGA	GCA	AGA	GAA	TCC	CAG	GAC	AGA	AAG	ATC	5399
Asn	His	Gln	Gly	Pro	Lys	Arg	Ala	Arg	Glu	Ser	Gln	Asp	Arg	Lys	Ile	
	1745				1750					1755					1760	
TTC	AGG	GGG	CTA	GAA	ATC	TGT	TGC	TAT	GGG	CCC	TTC	ACC	AAC	ATG	CCC	5447
Phe	Arg	Gly	Leu	Glu	Ile	Cys	Cys	Tyr	Gly	Pro	Phe	Thr	Asn	Met	Pro	
				1765					1770					1775		
ACA	GAT	CAA	CTG	GAA	TGG	ATG	GTA	CAG	CTG	TGT	GGT	GCT	TCT	GTG	GTG	5495

A000310

5,753,441

81

82

-continued

Thr	Asp	Gln	Leu	Glu	Trp	Met	Val	Gln	Leu	Cys	Gly	Ala	Ser	Val	Val	
			1780					1785					1790			
AAG	GAG	CTT	TCA	TCA	TTC	ACC	CTT	GGC	ACA	GGT	GTC	CAC	CCA	ATT	GTG	5543
Lys	Glu	Leu	Ser	Ser	Phe	Thr	Leu	Gly	Thr	Gly	Val	His	Pro	Ile	Val	
		1795					1800					1805				
GTT	GTG	CAG	CCA	GAT	GCC	TGG	ACA	GAG	GAC	AAT	GGC	TTC	CAT	GCA	ATT	5591
Val	Val	Gln	Pro	Asp	Ala	Trp	Thr	Glu	Asp	Asn	Gly	Phe	His	Ala	Ile	
		1810					1815					1820				
GGG	CAG	ATG	TGT	GAG	GCA	CCT	GTG	GTG	ACC	CGA	GAG	TGG	GTG	ITG	GAC	5639
Gly	Gln	Met	Cys	Glu	Tyr	Ala	Pro	Val	Val	Thr	Arg	Glu	Trp	Val	Leu	
					1830						1835				1840	
AGT	GTA	GCA	CTC	TAC	CAG	TGC	CAG	GAG	CTG	GAC	ACC	TAC	CTG	ATA	CCC	5687
Ser	Val	Ala	Leu	Tyr	Gln	Cys	Gln	Glu	Leu	Asp	Thr	Tyr	Leu	Ile	Pro	
				1845					1850					1855		
CAG	ATC	CCC	CAC	AGC	CAC	TAC	TGA	CTGCAGCCAG	CCACAGGTAC	AGAGCCACAG						5741
Gln	Ile	Pro	His	Ser	His	Tyr										
			1860													
GACCCCAAGA	ATGAGCTTAC	AAAGTGGCCT	TTCCAGGCCC	TGGGAGCTCC	TCTCACTCTT											5801
CAGTCCTTCT	ACTGTCCTGG	CTACTAAATA	TTTTATGTAC	ATCAGCCTGA	AAAGGACTTC											5861
TGGCTATGCA	AGGGTCCCTT	AAAGATTTTC	TGCTTGAAAGT	CTCCCTTGGA	AAT											5914

(2) INFORMATION FOR SEQ ID NO:2:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1863 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met	Asp	Leu	Ser	Ala	Leu	Arg	Val	Glu	Glu	Val	Gln	Asn	Val	Ile	Asn	
1				5					10					15		
Ala	Met	Gln	Lys	Ile	Leu	Glu	Cys	Pro	Ile	Cys	Leu	Glu	Leu	Ile	Lys	
		20					25					30				
Glu	Pro	Val	Ser	Thr	Lys	Cys	Asp	His	Ile	Phe	Cys	Lys	Phe	Cys	Met	
		35				40					45					
Leu	Lys	Leu	Leu	Asn	Gln	Lys	Lys	Gly	Pro	Ser	Gln	Cys	Pro	Leu	Cys	
		50			55						60					
Lys	Asn	Asp	Ile	Thr	Lys	Arg	Ser	Leu	Gln	Glu	Ser	Thr	Arg	Phe	Ser	
		65			70					75					80	
Gln	Leu	Val	Glu	Glu	Leu	Leu	Lys	Ile	Ile	Cys	Ala	Phe	Gln	Leu	Asp	
			85					90						95		
Thr	Gly	Leu	Glu	Tyr	Ala	Asn	Ser	Tyr	Asn	Phe	Ala	Lys	Lys	Glu	Asn	
		100					105						110			
Asn	Ser	Pro	Glu	His	Leu	Lys	Asp	Glu	Val	Ser	Ile	Ile	Gln	Ser	Met	
		115				120						125				
Gly	Tyr	Arg	Asn	Arg	Ala	Lys	Arg	Leu	Leu	Gln	Ser	Glu	Pro	Glu	Asn	
		130				135					140					
Pro	Ser	Leu	Gln	Glu	Thr	Ser	Leu	Ser	Val	Gln	Leu	Ser	Asn	Leu	Gly	
		145			150					155					160	
Thr	Val	Arg	Thr	Leu	Arg	Thr	Lys	Gln	Arg	Ile	Gln	Pro	Gln	Lys	Thr	
			165					170						175		
Ser	Val	Tyr	Ile	Glu	Leu	Gly	Ser	Asp	Ser	Ser	Glu	Asp	Thr	Val	Asn	
		180					185						190			
Lys	Ala	Thr	Tyr	Cys	Ser	Val	Gly	Asp	Gln	Glu	Leu	Leu	Gln	Ile	Thr	
		195					200					205				

5,753,441

83

84

-continued

Pro	Gln	Gly	Thr	Arg	Asp	Glu	Ile	Ser	Leu	Asp	Ser	Ala	Lys	Lys	Ala
	210					215					220				
Ala	Cys	Glu	Phe	Ser	Glu	Thr	Asp	Val	Thr	Asn	Thr	Glu	His	His	Gln
225					230					235					240
Pro	Ser	Asn	Asn	Asp	Leu	Asn	Thr	Thr	Glu	Lys	Arg	Ala	Ala	Glu	Arg
				245					250					255	
His	Pro	Glu	Lys	Tyr	Gln	Gly	Ser	Ser	Val	Ser	Asn	Leu	His	Val	Glu
			260					265					270		
Pro	Cys	Gly	Thr	Asn	Thr	His	Ala	Ser	Ser	Leu	Gln	His	Glu	Asn	Ser
		275					280					285			
Ser	Leu	Leu	Leu	Thr	Lys	Asp	Arg	Met	Asn	Val	Glu	Lys	Ala	Glu	Phe
	290					295					300				
Cys	Asn	Lys	Ser	Lys	Gln	Pro	Gly	Leu	Ala	Arg	Ser	Gln	His	Asn	Arg
305					310					315					320
Trp	Ala	Gly	Ser	Lys	Glu	Thr	Cys	Asn	Asp	Arg	Arg	Thr	Pro	Ser	Thr
				325					330					335	
Glu	Lys	Lys	Val	Asp	Leu	Asn	Ala	Asp	Pro	Leu	Cys	Glu	Arg	Lys	Glu
			340					345					350		
Trp	Asn	Lys	Gln	Lys	Leu	Pro	Cys	Ser	Glu	Asn	Pro	Arg	Asp	Thr	Glu
		355					360					365			
Asp	Val	Pro	Trp	Ile	Thr	Leu	Asn	Ser	Ser	Ile	Gln	Lys	Val	Asn	Glu
	370					375					380				
Trp	Phe	Ser	Arg	Ser	Asp	Glu	Leu	Leu	Gly	Ser	Asp	Asp	Ser	His	Asp
385					390					395					400
Gly	Glu	Ser	Glu	Ser	Asn	Ala	Lys	Val	Ala	Asp	Val	Leu	Asp	Val	Leu
				405					410					415	
Asn	Glu	Val	Asp	Glu	Tyr	Ser	Gly	Ser	Ser	Glu	Lys	Ile	Asp	Leu	Leu
			420					425					430		
Ala	Ser	Asp	Pro	His	Glu	Ala	Leu	Ile	Cys	Lys	Ser	Glu	Arg	Val	His
		435					440					445			
Ser	Lys	Ser	Val	Glu	Ser	Asn	Ile	Glu	Asp	Lys	Ile	Phe	Gly	Lys	Thr
	450					455					460				
Tyr	Arg	Lys	Lys	Ala	Ser	Leu	Pro	Asn	Leu	Ser	His	Val	Thr	Glu	Asn
465					470					475					480
Leu	Ile	Ile	Gly	Ala	Phe	Val	Thr	Glu	Pro	Gln	Ile	Ile	Gln	Glu	Arg
				485					490					495	
Pro	Leu	Thr	Asn	Lys	Leu	Lys	Arg	Lys	Arg	Arg	Pro	Thr	Ser	Gly	Leu
			500					505					510		
His	Pro	Glu	Asp	Phe	Ile	Lys	Lys	Ala	Asp	Leu	Ala	Val	Gln	Lys	Thr
		515					520					525			
Pro	Glu	Met	Ile	Asn	Gln	Gly	Thr	Asn	Gln	Thr	Glu	Gln	Asn	Gly	Gln
		530				535					540				
Val	Met	Asn	Ile	Thr	Asn	Ser	Gly	His	Glu	Asn	Lys	Thr	Lys	Gly	Asp
545					550					555					560
Ser	Ile	Gln	Asn	Glu	Lys	Asn	Pro	Asn	Pro	Ile	Glu	Ser	Leu	Glu	Lys
				565					570					575	
Glu	Ser	Ala	Phe	Lys	Thr	Lys	Ala	Glu	Pro	Ile	Ser	Ser	Ser	Ile	Ser
			580					585					590		
Asn	Met	Glu	Leu	Glu	Leu	Asn	Ile	His	Asn	Ser	Lys	Ala	Pro	Lys	Lys
		595					600					605			
Asn	Arg	Leu	Arg	Arg	Lys	Ser	Ser	Thr	Arg	His	Ile	His	Ala	Leu	Glu
	610					615					620				
Leu	Val	Val	Ser	Arg	Asn	Leu	Ser	Pro	Pro	Asn	Cys	Thr	Glu	Leu	Gln
625					630					635					640

A000312

5,753,441

85

86

-continued

Ile	Asp	Ser	Cys	Ser	Ser	Ser	Glu	Glu	Ile	Lys	Lys	Lys	Lys	Tyr	Asn
				645					650					655	
Gln	Met	Pro	Val	Arg	His	Ser	Arg	Asn	Leu	Gln	Leu	Met	Glu	Gly	Lys
			660					665					670		
Glu	Pro	Ala	Thr	Gly	Ala	Lys	Lys	Ser	Asn	Lys	Pro	Asn	Glu	Gln	Thr
		675					680					685			
Ser	Lys	Arg	His	Asp	Ser	Asp	Thr	Phe	Pro	Glu	Leu	Lys	Leu	Thr	Asn
	690					695					700				
Ala	Pro	Gly	Ser	Phe	Thr	Lys	Cys	Ser	Asn	Thr	Ser	Glu	Leu	Lys	Glu
	705				710					715					720
Phe	Val	Asn	Pro	Ser	Leu	Pro	Arg	Glu	Glu	Lys	Glu	Glu	Lys	Leu	Glu
				725				730						735	
Thr	Val	Lys	Val	Ser	Asn	Asn	Ala	Glu	Asp	Pro	Lys	Asp	Leu	Met	Leu
			740					745					750		
Ser	Gly	Glu	Arg	Val	Leu	Gln	Thr	Glu	Arg	Ser	Val	Glu	Ser	Ser	Ser
		755					760					765			
Ile	Ser	Leu	Val	Pro	Gly	Thr	Asp	Tyr	Gly	Thr	Gln	Glu	Ser	Ile	Ser
	770					775					780				
Leu	Leu	Glu	Val	Ser	Thr	Leu	Gly	Lys	Ala	Lys	Thr	Glu	Pro	Asn	Lys
	785				790					795					800
Cys	Val	Ser	Gln	Cys	Ala	Ala	Phe	Glu	Asn	Pro	Lys	Gly	Leu	Ile	His
				805					810					815	
Gly	Cys	Ser	Lys	Asp	Asn	Arg	Asn	Asp	Thr	Glu	Gly	Phe	Lys	Tyr	Pro
			820					825					830		
Leu	Gly	His	Glu	Val	Asn	His	Ser	Arg	Glu	Thr	Ser	Ile	Glu	Met	Glu
		835					840					845			
Glu	Ser	Glu	Leu	Asp	Ala	Gln	Tyr	Leu	Gln	Asn	Thr	Phe	Lys	Val	Ser
	850					855					860				
Lys	Arg	Gln	Ser	Phe	Ala	Pro	Phe	Ser	Asn	Pro	Gly	Asn	Ala	Glu	Glu
	865				870					875					880
Glu	Cys	Ala	Thr	Phe	Ser	Ala	His	Ser	Gly	Ser	Leu	Lys	Lys	Gln	Ser
				885					890					895	
Pro	Lys	Val	Thr	Phe	Glu	Cys	Glu	Gln	Lys	Glu	Glu	Asn	Gln	Gly	Lys
			900					905					910		
Asn	Glu	Ser	Asn	Ile	Lys	Pro	Val	Gln	Thr	Val	Asn	Ile	Thr	Ala	Gly
		915					920					925			
Phe	Pro	Val	Val	Gly	Gln	Lys	Asp	Lys	Pro	Val	Asp	Asn	Ala	Lys	Cys
	930					935					940				
Ser	Ile	Lys	Gly	Gly	Ser	Arg	Phe	Cys	Leu	Ser	Ser	Gln	Phe	Arg	Gly
	945				950					955					960
Asn	Glu	Thr	Gly	Leu	Ile	Thr	Pro	Asn	Lys	His	Gly	Leu	Leu	Gln	Asn
				965					970					975	
Pro	Tyr	Arg	Ile	Pro	Pro	Leu	Phe	Pro	Ile	Lys	Ser	Phe	Val	Lys	Thr
			980					985					990		
Lys	Cys	Lys	Lys	Asn	Leu	Leu	Glu	Glu	Asn	Phe	Glu	Glu	His	Ser	Met
		995					1000					1005			
Ser	Pro	Glu	Arg	Glu	Met	Gly	Asn	Glu	Asn	Ile	Pro	Ser	Thr	Val	Ser
	1010					1015					1020				
Thr	Ile	Ser	Arg	Asn	Asn	Ile	Arg	Glu	Asn	Val	Phe	Lys	Glu	Ala	Ser
	1025				1030					1035					1040
Ser	Ser	Asn	Ile	Asn	Glu	Val	Gly	Ser	Ser	Thr	Asn	Glu	Val	Gly	Ser
				1045				1050						1055	
Ser	Ile	Asn	Glu	Ile	Gly	Ser	Ser	Asp	Glu	Asn	Ile	Gln	Ala	Glu	Leu

A000313

5,753,441

87

88

-continued

1060										1065										1070									
Gly	Arg	Asn	Arg	Gly	Pro	Lys	Leu	Asn	Ala	Met	Leu	Arg	Leu	Gly	Val														
		1075					1080					1085																	
Leu	Gln	Pro	Glu	Val	Tyr	Lys	Gln	Ser	Leu	Pro	Gly	Ser	Asn	Cys	Lys														
	1090					1095					1100																		
His	Pro	Glu	Ile	Lys	Lys	Gln	Glu	Tyr	Glu	Glu	Val	Val	Gln	Thr	Val														
	1105				1110					1115					1120														
Asn	Thr	Asp	Phe	Ser	Pro	Tyr	Leu	Ile	Ser	Asp	Asa	Leu	Glu	Gln	Pro														
			1125						1130					1135															
Met	Gly	Ser	Ser	His	Ala	Ser	Gln	Val	Cys	Ser	Glu	Thr	Pro	Asp	Asp														
			1140					1145					1150																
Leu	Leu	Asp	Asp	Gly	Glu	Ile	Lys	Glu	Asp	Thr	Ser	Phe	Ala	Glu	Asn														
		1155					1160					1165																	
Asp	Ile	Lys	Glu	Ser	Ser	Ala	Val	Phe	Ser	Lys	Ser	Val	Gln	Lys	Gly														
	1170					1175					1180																		
Glu	Leu	Ser	Arg	Ser	Pro	Ser	Pro	Phe	Thr	His	Thr	His	Leu	Ala	Gln														
	1185				1190					1195					1200														
Gly	Tyr	Arg	Arg	Gly	Ala	Lys	Lys	Leu	Glu	Ser	Ser	Glu	Glu	Asa	Leu														
			1205						1210					1215															
Ser	Ser	Glu	Asp	Glu	Glu	Leu	Pro	Cys	Phe	Gln	His	Leu	Leu	Phe	Gly														
		1220					1225						1230																
Lys	Val	Asn	Asn	Ile	Pro	Ser	Gln	Ser	Thr	Arg	His	Ser	Thr	Val	Ala														
		1235					1240					1245																	
Thr	Glu	Cys	Leu	Ser	Lys	Asn	Thr	Glu	Glu	Asn	Leu	Leu	Ser	Leu	Lys														
	1250					1255					1260																		
Asn	Ser	Leu	Asn	Asp	Cys	Ser	Asn	Gln	Val	Ile	Leu	Ala	Lys	Ala	Ser														
	1265				1270				1275				1280																
Gln	Glu	His	His	Leu	Ser	Glu	Glu	Thr	Lys	Cys	Ser	Ala	Ser	Leu	Phe														
			1285						1290				1295																
Ser	Ser	Gln	Cys	Ser	Glu	Leu	Glu	Asp	Leu	Thr	Ala	Asn	Thr	Asn	Thr														
		1300					1305						1310																
Gln	Asp	Pro	Phe	Leu	Ile	Gly	Ser	Ser	Lys	Gln	Met	Arg	His	Gln	Ser														
		1315					1320					1325																	
Glu	Ser	Gln	Gly	Val	Gly	Leu	Ser	Asp	Lys	Glu	Leu	Val	Ser	Asp	Asp														
	1330					1335					1340																		
Glu	Glu	Arg	Gly	Thr	Gly	Leu	Glu	Glu	Asn	Asn	Gln	Glu	Glu	Gln	Ser														
	1345				1350				1355				1360																
Met	Asp	Ser	Asn	Leu	Gly	Glu	Ala	Ala	Ser	Gly	Cys	Glu	Ser	Glu	Thr														
			1365						1370				1375																
Ser	Val	Ser	Glu	Asp	Cys	Ser	Gly	Leu	Ser	Ser	Gln	Ser	Asp	Ile	Leu														
		1380					1385						1390																
Thr	Thr	Gln	Gln	Arg	Asp	Thr	Met	Gln	His	Asn	Leu	Ile	Lys	Leu	Gln														
		1395					1400					1405																	
Gln	Glu	Met	Ala	Glu	Leu	Glu	Ala	Val	Leu	Glu	Gln	His	Gly	Ser	Gln														
	1410					1415					1420																		
Pro	Ser	Asn	Ser	Tyr	Pro	Ser	Ile	Ile	Ser	Asp	Ser	Ser	Ala	Leu	Glu														
	1425				1430				1435				1440																
Asp	Leu	Arg	Asa	Pro	Glu	Gln	Ser	Thr	Ser	Glu	Lys	Ala	Val	Leu	Thr														
			1445						1450				1455																
Ser	Gln	Lys	Ser	Ser	Glu	Tyr	Pro	Ile	Ser	Gln	Asn	Pro	Glu	Gly	Leu														
		1460					1465						1470																
Ser	Ala	Asp	Lys	Phe	Glu	Val	Ser	Ala	Asp	Ser	Ser	Thr	Ser	Lys	Asn														
	1475						1480					1485																	

A000314

5,753,441

89

90

-continued

Lys	Glu	Pro	Gly	Val	Glu	Arg	Ser	Ser	Pro	Ser	Lys	Cys	Pro	Ser	Leu
1490						1495					1500				
Asp	Asp	Arg	Trp	Tyr	Met	His	Ser	Cys	Ser	Gly	Ser	Leu	Gln	Asn	Arg
1505					1510					1515					1520
Asn	Tyr	Pro	Ser	Gln	Glu	Glu	Leu	Ile	Lys	Val	Val	Asp	Val	Glu	Glu
				1525					1530					1535	
Gln	Gln	Leu	Glu	Glu	Ser	Gly	Pro	His	Asp	Leu	Thr	Glu	Thr	Ser	Tyr
			1540					1545					1550		
Leu	Pro	Arg	Gln	Asp	Leu	Glu	Gly	Thr	Pro	Tyr	Leu	Glu	Ser	Gly	Ile
		1555					1560					1565			
Ser	Leu	Phe	Ser	Asp	Asp	Pro	Glu	Ser	Asp	Pro	Ser	Glu	Asp	Arg	Ala
	1570					1575					1580				
Pro	Glu	Ser	Ala	Arg	Val	Gly	Asn	Ile	Pro	Ser	Ser	Thr	Ser	Ala	Leu
1585					1590					1595					1600
Lys	Val	Pro	Gln	Leu	Lys	Val	Ala	Glu	Ser	Ala	Gln	Ser	Pro	Ala	Ala
				1605				1610						1615	
Ala	His	Thr	Thr	Asp	Thr	Ala	Gly	Tyr	Asn	Ala	Met	Glu	Glu	Ser	Val
			1620					1625					1630		
Ser	Arg	Glu	Lys	Pro	Glu	Leu	Thr	Ala	Ser	Thr	Glu	Arg	Val	Asn	Lys
		1635					1640					1645			
Arg	Met	Ser	Met	Val	Val	Ser	Gly	Leu	Thr	Pro	Glu	Glu	Phe	Met	Leu
	1650					1655					1660				
Val	Tyr	Lys	Phe	Ala	Arg	Lys	His	His	Ile	Thr	Leu	Thr	Asn	Leu	Ile
1665					1670					1675					1680
Thr	Glu	Glu	Thr	Thr	His	Val	Val	Met	Lys	Thr	Asp	Ala	Glu	Phe	Val
				1685				1690						1695	
Cys	Glu	Arg	Thr	Leu	Lys	Tyr	Phe	Leu	Gly	Ile	Ala	Gly	Gly	Lys	Trp
			1700					1705					1710		
Val	Val	Ser	Tyr	Phe	Trp	Val	Thr	Gln	Ser	Ile	Lys	Glu	Arg	Lys	Met
		1715					1720					1725			
Leu	Asn	Glu	His	Asp	Phe	Glu	Val	Arg	Gly	Asp	Val	Val	Asn	Gly	Arg
	1730					1735					1740				
Asn	His	Gln	Gly	Pro	Lys	Arg	Ala	Arg	Glu	Ser	Gln	Asp	Arg	Lys	Ile
1745					1750					1755					1760
Phe	Arg	Gly	Leu	Glu	Ile	Cys	Cys	Tyr	Gly	Pro	Phe	Thr	Asn	Met	Pro
			1765						1770					1775	
Thr	Asp	Gln	Leu	Glu	Trp	Met	Val	Gln	Leu	Cys	Gly	Ala	Ser	Val	Val
			1780					1785					1790		
Lys	Glu	Leu	Ser	Ser	Phe	Thr	Leu	Gly	Thr	Gly	Val	His	Pro	Ile	Val
		1795					1800					1805			
Val	Val	Gln	Pro	Asp	Ala	Trp	Thr	Glu	Asp	Asn	Gly	Phe	His	Ala	Ile
		1810				1815					1820				
Gly	Gln	Met	Cys	Glu	Ala	Pro	Val	Val	Thr	Arg	Glu	Trp	Val	Leu	Asp
1825					1830					1835					1840
Ser	Val	Ala	Leu	Tyr	Gln	Cys	Gln	Glu	Leu	Asp	Thr	Tyr	Leu	Ile	Pro
			1845					1850						1855	
Gln	Ile	Pro	His	Ser	His	Tyr									
			1860												

(2) INFORMATION FOR SEQ ID NO:3:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

A000315

5,753,441

91

92

-continued

(i i) MOLECULE TYPE: DNA (genomic)
(i i i) HYPOTHETICAL: NO
(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens
(v i i) IMMEDIATE SOURCE:
(B) CLONE: s754 A
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:3:
C T A G C C T G G G C A A C A A A C G A 2 0
(2) INFORMATION FOR SEQ ID NO:4:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(i i) MOLECULE TYPE: DNA (genomic)
(i i i) HYPOTHETICAL: NO
(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens
(v i i) IMMEDIATE SOURCE:
(B) CLONE: s754 B
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:4:
G C A G G A A G C A G G A A T G G A A C 2 0
(2) INFORMATION FOR SEQ ID NO:5:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(i i) MOLECULE TYPE: DNA (genomic)
(i i i) HYPOTHETICAL: NO
(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens
(v i i) IMMEDIATE SOURCE:
(B) CLONE: s975 A
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:5:
T A G G A G A T G G A T T A T T G G T G 2 0
(2) INFORMATION FOR SEQ ID NO:6:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(i i) MOLECULE TYPE: DNA (genomic)
(i i i) HYPOTHETICAL: NO
(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens
(v i i) IMMEDIATE SOURCE:
(B) CLONE: s975 B

A000316

5,753,441

93

94

-continued

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:6:
AGGCAACTTT GCAATGAGTG 2 0

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(v i i) IMMEDIATE SOURCE:
(B) CLONE: tdj1474 A

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:7:
CAGAGTGAGA CCTGTCTCA AA 2 2

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(v i i) IMMEDIATE SOURCE:
(B) CLONE: tdj1474 B

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:8:
TTCTGCAAAC ACCTTAACT CAG 2 3

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(v i i) IMMEDIATE SOURCE:
(B) CLONE: tdj1239 A

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:9:
AACCTGGAAG GCAGAGGTTG 2 0

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single

5,753,441

95

96

-continued

(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens

(v i i) IMMEDIATE SOURCE:

(B) CLONE: rdj1239 B

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:10:

TCTGTACCTGCTAAGCAGTGG21

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 111 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: cDNA

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens

(i x) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 2..111

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:11:

G GKC TTA CTC TGT TGT CCC AGC TGG AGT ACA GWG TGC GAT CAT GAG46

Xaa Leu Leu Cys Cys Pro Ser Trp Ser Thr Xaa Cys Asp His Glu

186518701875

GCT TAC TGT TGC TTG ACT CCT AGG CTC AAG CGA TCC TAT CAC CTC AGT94

Ala Tyr Cys Cys Leu Thr Pro Arg Leu Lys Arg Ser Tyr His Leu Ser

1880188518901895

CTC CAA GTA GCT GGA CT111

Leu Gln Val Ala Gly

1900

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 36 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: protein

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:12:

Xaa Leu Leu Cys Cys Pro Ser Trp Ser Thr Xaa Cys Asp His Glu Ala15

15

Tyr Cys Cys Leu Thr Pro Arg Leu Lys Arg Ser Tyr His Leu Ser Leu202530

Gln Val Ala Gly35

35

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1534 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

5,753,441

97

98

-continued

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(i v) ANTI-SENSE: NO

(v i) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:13:

GAGGCTAGAG	GGCAGGCACT	TTATGGCAAA	CTCAGGTAGA	ATTCTTCCTC	ITCCGTCTCT	60
TTCCTTTTAC	GTTCATCGGG	AGACTGGGTG	GCAATCGCAG	CCCGAGAGAC	GCATGGCTCT	120
TTCTGCCCTC	CATCCTCTGA	TGTACCTTGA	TTTCGTATT	TGAGAGGCTG	CTGCTTAGCG	180
GTAGCCCTT	GGTTTCCGTG	GCAACGGAAA	AGCGCGGGAA	TTACAGATAA	ATTAAAACTG	240
CGACTGCGCG	GCCTGAGCTC	GCTGAGACTT	CCTGGACCCC	GCACCAGGCT	GTGGGGTTTC	300
TCAGATAACT	GGGCCCCCTG	GCTCAGGAGG	CCTTCACCCT	CTGCTCTGGG	TAAAGGTAGT	360
AGAGTCCCG	GAAAGGGACA	GGGGGCCCAA	GTGATGCTCT	GGGGTACTGG	CGTGGGAGAG	420
TGGATTTCCG	AAGCTGACAG	ATGGGTATT	TTTGACGGGG	GGTAGGGGCG	GAACCTGAGA	480
GGCGTAAGGC	GTTGTGAACC	CTGGGGAGGG	GGGCAGTTTG	TAGGTCGCGA	GGGAAGCGCT	540
GAGGATCAGG	AAGGGGGCAC	TGAGTGTCCG	TGGGGGAATC	CTCGTGATAG	GAAC TGGAAT	600
ATGCCTTGAG	GGGGACACTA	TGCTTTTAAA	AACGTCGGCT	GGTCATGAGG	TCAGGAGTTC	660
CAGACCAGCC	TGACCAACGT	GGTGAAACTC	CGTCTCTACT	AAAAATACNA	AAATTAGCCG	720
GGCGTGGTGC	CGCTCCAGCT	ACTCAGGAGG	CTGAGGCAGG	AGAATCGCTA	GAACCCGGGA	780
GGCGGAGGTT	GCAGTGAGCC	GAGATCGCGC	CATTGCACTC	CAGCCTGGGC	GACAGAGCGA	840
GACTGTCTCA	AAACAAAACA	AAACAAAACA	AAACAAAAAA	CACCGGCTGG	TATGTATGAG	900
AGGATGGGAC	CTTGTGGAAG	AAGAGGTGCC	AGGAATATGT	CTGGGAAGGG	GAGGAGACAG	960
GATTTTGTGG	GAGGGAGAAC	TTAAGAACTG	GATCCATTTG	CGCCATTGAG	AAAACGCAAG	1020
AGGGAAGTAG	AGGAGCGTCA	GTAGTAACAG	ATGCTGCCGG	CAGGGATGTG	CTTGAGGAGG	1080
ATCCAGAGAT	GAGAGCAGGT	CACTGGGAAA	GGTTAGGGGC	GGGGAGGCCT	TGATTGGTGT	1140
TGGTTTGGTC	GTTGTTGATT	TTGGTTTTAT	GCAAGAAAAA	GAAAAACAAC	AGAAACATTG	1200
GAGAAAGCTA	AGGCTACCAC	CACCTACCCG	GTCAGTCACT	CCTCTGTAGC	TTTCTCTTTC	1260
TTGGAGAAAAG	GAAAAAGACCC	AAGGGGTTGG	CAOCGATATG	TGAAAAAATT	CAGAATTTAT	1320
GTTGTCTAAT	TACAAAAAGC	AACTTCTAGA	ATCTTTAAAA	ATAAAGGACG	TTGTCATTAG	1380
TTCTTCTGGT	TTGTATTATT	CTAAAACCTT	CCAAATCTTC	AAATTTACTT	TATTTTAAAA	1440
TGATAAAATG	AAGTTGTCAT	TTTATAAACC	TTTTAAAAAG	ATATATATAT	ATGTTTTTCT	1500
AAATGTATAA	AGTTCATTGG	AACAGAAAGA	AAATG			1532

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1924 base pairs

(R) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(i v) ANTI-SENSE: NO

(v i) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens

5,753,441

99

100

-continued

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:14:

GAGGCTAGAG	GGCAGGCACT	TTATGGCAAA	CTCAGGTAGA	ATTCTTCTCT	TTCCGTCCTC	60
TTCCTTTTAC	GTCAATCGGG	AGACTGGGTG	GCAATCGCAG	CCCAGAGAGAC	GCATGGCTCT	120
TTCTGCCCTC	CATCCTCTGA	TGTACCTTGA	TTTCGTATT	TGAGAGGCTG	CTGCTTAGCG	180
GTAGCCCCCT	GGTTTCCGTG	GCAACGGAAA	AGCGCGGGAA	TTACAGATAA	ATTAAAACTG	240
CGACTGCGCG	GCCTGAGCTC	GCTGAGACTT	CCTGGACCCC	GCACCAGGCT	GTGGGGTTTC	300
TCAGATAACT	GGGCCCCCTG	GCTCAGGAGG	CCTTCACCTT	CTGCTCTGGG	TAAAGGTAGT	360
AGAGTCCCG	GAAAGGGACA	GGGGGCCCAA	GTGATGCTCT	GGGGTACTGG	CGTGGGAGAG	420
TGGATTTCG	AACTGACAG	ATGGGTATT	TTTGACGGG	GGTAGGGGCG	GAACCTGAGA	480
GGCGTAAGGC	GTTGTGAACC	CTGGGGAGGG	GGGCAGTTTG	TAGGTCGCGA	GGGAAGCGCT	540
GAGGATCAGG	AAGGGGGCAC	TGAGTGTCCG	TGGGGGAATC	CTCGTGATAG	GAACCTGAAT	600
ATGCCTTGAG	GGGGACACTA	TGTCTTTAAA	AACGTCGGCT	GGTCATGAGG	TCAGGAGTTC	660
CAGACCAGCC	TGACCAACGT	GGTGAAACTC	CGTCTCTACT	AAAAATACNA	AAATTAGCCG	720
GGCGTGGTGC	CGCTCCAGCT	ACTCAGGAGG	CTGAGGCAGG	AGAATCGCTA	GAACCCGGGA	780
GGCGGAGGTT	GCAGTGAGCC	GAGATCGCGC	CATTGCACTC	CAGCCTGGGC	GACAGAGCGA	840
GACTGTCTCA	AAACAAAACA	AAACAAAACA	AAACAAAAAA	CACCGGCTGG	TATGTATGAG	900
AGGATGGGAC	CTTGTGGAAG	AAGAGGTGCC	AGGAATATGT	CTGGGAAGGG	GAGGAGACAG	960
GATTTTGTGG	GAGGGAGAAC	TTAAGAACTG	GATCCATTTG	CGCCATTGAG	AAAGCGCAAG	1020
AGGGAAGTAG	AGGAGCGTCA	GTAATAACAG	ATGCTGCCGG	CAGGGATGTG	CTTGAGGAGG	1080
ATCCAGAGAT	GAGAGCAGGT	CACTGGGAAA	GGTTAGGGGC	GGGGAGGCCT	TGATTGGTGT	1140
TGGTTTGGTC	GTGTGTGATT	TTGGTTTTAT	GCAAGAAAAA	GAAAACAACC	AGAAACATTG	1200
GAGAAAAGCTA	AGGCTACCAC	CACCTACCCG	GTCACTACT	CCTCTGTAGC	TTTCTCTTTC	1260
TTGGAGAAAG	GAAAAGACCC	AAGGGGTTGG	CAGCGATATG	TGAAAAAATT	CAGAATTTAT	1320
TTTGTCTAAT	TACAAAAAGC	AACCTCTAGA	ATCTTTAAAA	ATAAAGGACG	TTGTCATTAG	1380
TTCTTCTGGT	TTGTATTATT	CTAAACCTT	CCAAATCTTC	AAATTTACTT	TATTTTAAAA	1440
TGATAAAATG	AAGTTGTCA	TTTATAAACC	TTTTAAAAAG	ATATATATAT	ATGTTTTTCT	1500
AATGTGTTAA	AGTTCATTGG	AACAGAAAGA	AATGGATTTA	TCTGCTCTTC	GCCTTGAAGA	1560
AGTACAAAAT	GTCATTAATG	CTATGCAGAA	AATCTTAGAG	TGTCCTCTCT	GGTAAGTCAG	1620
CACAAGAGTG	TATTAATTGG	GGATTCCTAT	GATTATCTCT	TATGCAAAATG	AACAGAATTG	1680
ACCTTACATA	CTAGGGAAAG	AAAGACATGT	CTAGTAAGAT	TAGGCTATTG	TAATTGCTGA	1740
TTTTCTTAAC	TGAAGAACTT	TAAAAATATA	GAAAATGATT	CCTTGTTCTC	CATCCACTCT	1800
GCCTCTCCCA	CTCCTCTCCT	TTTCAACACA	ATCCTGTGGT	CCGGGAAAGA	CAGGGCTCTG	1860
TCTTGATTGG	TTCTGCACTG	GGCAGGATCT	GTTAGATACT	GCATTTGCTT	TCTCCAGCTC	1920
TAAA						1920

(2) INFORMATION FOR SEO ID NO:15:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 631 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

5,753,441

101

102

-continued

(i i i) HYPOTHETICAL: NO

(i v) ANTI-SENSE: NO

(v i) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:15:

A A A T G C T G A T	G A T A G T A T A G	A G T A T T G A A G	G G A T C A A T A T	A A T T C T G T T T	T G A T A T C T G A	60
A A G C T C A C T G	A A G G T A A G G A	T C G T A T T C T C	T G C T G T A T T C	T C A G T T C C T G	A C A C A G C A G A	120
C A T T T A A T A A	A T A T T G A A C G	A A C T T G A G G C	C T T A T G T T G A	C T C A G T C A T A	A C A G C T C A A A	180
G T T G A A C T T A	T T C A C T A A G A	A T A G C T T T A T	T T T T A A A T A A	A T T A T T G A G C	C T C A T T T A T T	240
T T C T T T T T C T	C C C C C C C C T A	C C C T G C T A G T	C T G G A G T T G A	T C A A G G A A C C	T G T C T C C A C A	300
A A G T G T G A C C	A C A T A T T T T G	C A A G T A A G T T	T G A A T G T G T T	A T G T G G C T C C	A T T A T T A G C T	360
T T T G T T T T T G	T C C T T C A T A A	C C C A G G A A A C	A C C T A A C T T T	A T A G A A G C T T	T A C T T T C T T C	420
A A T T A A G T G A	G A A C G A A A A T	C C A A C T C C A T	T T C A T T C T T T	C T C A G A G A G T	A T A T A G T T A T	480
C A A A A G T T G G	T T G T A A T C A T	A G T T C C T G G T	A A A G T T T T G A	C A T A T A T T A T	C T T T T T T T T T	540
T T T T G A G A C A	A G T C T C G C T C	T G T C G C C C A G	G C T G G A G T G C	A G T G G C A T G A	G G C T T G C T C A	600
C T G C A C C T C C	G C C C C C G A G T	T C A G C G A C T C	T			630

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 481 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(i v) ANTI-SENSE: NO

(v i) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:16:

TTGAGATCTAG	ACCACATGGT	CAAAGAGATA	GAATGTGAGC	AATAAATGAA	CCTTAAATTT	60
TTCAACAGCT	ACTTTTTTTT	TTTTTTTTTG	AGACAGGGKC	TTACTCTGTT	GTCCAGCTG	120
GAGTACAGWG	TGCGATCATG	AGGCTTACTG	TTGCTTGACT	CCTAGGCTCA	AGCGATCCTA	180
TCACCTCAGT	CTCCAAGTAG	CTGGACTGTA	AGTGACACC	ACCATATCCA	GCTAAATTTT	240
GTGTTTTCTG	TAGAGACGGG	GTTTCGCCAT	GTTTCCAGG	CTGGTCTTGA	ACTTTGGGCT	300
TAACCCGTCT	GCCCACCTAG	GCATCCCAA	GTGCTAGGAT	TACAGGTGTG	AGTCATCATG	360
CCTGGCCAGT	ATTTTAGTTA	GCCTGTCTT	TTCAAGTCAT	ATACAAGTTC	ATTTTCTTTT	420
AAGTTTAGTT	AACAACCTTA	TATCATGTAT	TCTTTTCTAG	CATAAAGAAA	GATTCGAGGC	480
C						481

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 522 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

5,753,441

103

104

-continued

(i i i) HYPOTHETICAL: NO

(i v) ANTI-SENSE: NO

(v i) ORIGINAL SOURCE:

(A) ORGANISM: *Homo sapiens*

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:17:

TGTGATCATA	ACAGTAAGCC	ATATGCATGT	AAGTTCAGTT	TTCATAGATC	ATTGCTTATG	60
TAGTTTAGGT	TTTTGCTTAT	GCAGCATCCA	AAAAACAATTA	GGAAACTATT	GCTTGTAATT	120
CACCTGCCAT	TACTTTTTAA	ATGGCTCTTA	AGGGCAGTIG	TGAGATTATC	TTTTCATGGC	180
TATTTGCCIT	TTGAGTATTC	TTTCTACAAA	AGGAAGTAAA	TTAAAATTGTT	CTTCTTTTCT	240
TTATAATTTA	TAGATTTTGC	ATGCTGAAAC	TTCTCAACCA	GAAGAAAGGG	CCTTCACAGT	300
GTCCTTTATG	TAAGAATGAT	ATAACCAAAA	GGTATATAAT	TTGGTAATGA	TGCTAGGTTG	360
GAAGCAACCA	CAGTAGGAAA	AAGTAGAAAT	TATTTAATAA	CATAGCGTTC	CTATAAAACC	420
ATTATCAGA	AAAAATTTATA	AAAGAGTTTT	TAGCACACAG	TAAATTATTT	CCAAAGTTAT	480
TTTCCTGAAA	GIITTAATGGG	CACTGCCCT	ATACAGGTAT	TG		522

(2) INFORMATION FOR SEO ID NO:18:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 465 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(i v) ANTI-SENSE: NO

(v i) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:18:

GGTAGGCTTA	AATGAATGAC	AAAAAGTTAC	TAAATCACTG	CCATCACACG	GTTTATACAG	60
ATGTCAATGA	TGTATTGATT	ATAGAGGTTT	TCTACTGTTG	CTGCATCTTA	TTTTTATTTG	120
TTTACATGTC	TTTTCTTATT	TTAGTGTCC	TAAAAGGTTG	ATAATCACIT	GCTGAGTGTG	180
TTTCTCAAAC	AATTTAATTT	CAGGAGCCTA	CAAGAAAGTA	CGAGATTTAG	TCAACTTGTT	240
GAAGAGCTAT	TGAAAAATCAT	TGTGCTTTT	CAGCTTGACA	CAGGTTTGG	GTGTAAGTGT	300
TGAATATCCC	AAGAATGACA	CTCAAGTGCT	GTCCATGAAA	ACTCAGGAAG	TTTGCACAAT	360
TACTTTCTAT	GACGTGGTGA	TAAGACCTTT	TAGTCTAGGT	TAATTTTAGT	TCTGTATCTG	420
TAATCTATTT	TAAAAAATTA	CTCCCACTGG	CTCACACCT	TATTT		465

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 513 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(i v) ANTI-SENSE: NO

(v i) ORIGINAL SOURCE:

(A) ORGANISM: *Homo sapiens*

5,753,441

105

106

-continued

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:19:

AAAAAATCAC	AGGTAACCTT	AATGCATTGT	CTTAACACAA	CAAAGAGCAT	ACATAGGGTT	60
TCTCTTGGTT	TCTTTGATTA	TAATTCATAC	ATTTTCTCT	AACTGCAAAC	ATAATGTTTT	120
CCCTTGATAT	TTACAGATGC	AAACAGCTAT	AATTTTGCAA	AAAAGGAAAA	TAACTCTCCT	180
GAACATCTAA	AAGATGAAGT	TTCTATCATC	CAAAGTATGG	GCTACAGAAA	CCGTGCCAAA	240
AGACTTCTAC	AGAGTGAACC	CGAAAATCCT	TCCTTGGTAA	AACCATIIGT	TTTCTTCTTC	300
TTCTTCTTCT	TCTTTTCTTT	TTTTTTTCTT	TTTTTTTTTG	AGATGGAGTC	TTGCTCTGTG	360
GCCCAGGCTA	GAAGCAGTCC	TCCTGCCCTA	GCCNCCTTAG	TAGCTGGGAT	TACAGGCACG	420
CGCACCATGC	CAGGCTAATT	TTTGTAATTT	TAGTAGAGAC	GGGGTTTCAT	CATGTTGGCC	480
AGGCTGGTCT	CGAACTCCTA	ACCTCAGGTG	ATC			513

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6769 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(i v) ANTI-SENSE: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:20:

ATGATGGAGA	TCTTAAAAAG	TAATCATTCT	GGGGCTGGGC	G TAGTAGCTT	GCACCTGTAA	60
TCCAGCACT	TCGGGAGGCT	GAGGCAGGCA	GATAATTTGA	GGTCAGGAGT	TTGAGACCAG	120
CCTGGCCAAC	ATGGTGAAAC	CCATCTCTAC	TAAAAATACA	AAAATTAGCT	GGGTGTGGTG	180
GCACGTACCT	GTAATCCCAg	CTACTCGGGA	GGCGGAAGGCA	CAAGAATTGC	TTGAACCTAG	240
GACGCGGAGG	TTGCAGCGAG	CCAAGATCGC	GCCACTGCAC	TCCAGCCTGG	GCCGTAGAGT	300
GAGACTCIGT	CTCAAAAAAG	AAAAAAAAGT	AATTGTTCTA	GCTGGGCGCA	GTGGCTCTTG	360
CCTGTAATCC	CAGCACTTTG	GGAGGCCAAG	GCGGGTGGAT	CTCGAGTCCT	AGAGTTCAAG	420
ACCAGCCTAG	GCAATGTGGT	GAAACCCCAT	CGCTACAAAA	AATACAAAAA	TTAGCCAGGC	480
ATGGTGGCGT	GCGCATGTAG	TCCAGCTCC	TTGGGAAGGCT	GAGGTGGGAG	GATCACTTGA	540
ACCCAGGAGA	CAGAGGTTGC	AGTGAACCGA	GATCACGCCA	CCACGCTCCA	GCCTGGGCAA	600
CAGAACAAGA	CTCTGTCTAA	AAAAATACAA	ATAAAATAAA	AGTAGTTCTC	ACAGTACCAG	660
CATTCATTTT	TCAAAAGATA	TAGAGCTAAA	AAGGAAGGAA	AAAAAAAAGTA	ATGTTGGGCT	720
TTTAAATACT	CGTTCCTATA	CTAAATGTTT	TTAGGAGTGC	TGGGGTTTTA	TTGTCATCAT	780
TTATCCTTTT	TAAAAATGTT	ATTGGCCAGG	CACGGTGGCT	CATGGCTGTA	ATCCCAAGCAC	840
TTTGGGAGGC	CGAGGCAGGC	AGATCACCTG	AGBTCAGGAG	TGTGAGACCA	GCCTGGGCAA	900
CATGGCGAAA	CCTGTCTCTA	CTAAAAATAC	AAAAATTAAc	TAGGCGTGGT	GGTGTACGCC	960
TGTAGTCCCA	GCTACTCGGG	AGGCTGAGGC	AGGAGAATCA	ACTGAACCAG	GGAGGTGGAG	1020
GTTGCAGTGT	GCCGAGATCA	CGCCACTGCA	CTCTAGCCTG	GCAACAGAGC	AAGATTCTGT	1080
CTCAAAAAAA	AAAAACATAT	ATACACATAT	ATCCCAAAAGT	GCTGGGATTA	CATATATATA	1140
TATATATATA	TATTATATAT	ATATATATAT	ATATATGTGA	TATATATGTG	ATATATATAT	1200

5,753,441	
107	108
-continued	
AACATATATA TATGTAATAT ATATGTGATA TATATATAAT ATATATAATG AATATATATG	1260
TGATATATAT ATATACACAC ACACACACAT ATATATGTAT GTGTGTGTAC ACACACACAC	1320
ACAAATTAGC CAGGCATAGT TGCACACGCT TGGTAGACCC AGCTACTCAG GAGGCTGAGG	1380
GAGGAGAATC TCTTGAACIT AGGAGGCGGA GGTTCAGTG AGCTGAGATT GCGCCACTGC	1440
ACTCCAGCCT GGGTGACAGA GCAGGACTCT GTACACCCCC CAAAACAAAA AAAAAAGTTA	1500
TCAGATGTGA TTGGAATGTA TATCAAGTAT CAGCTTCAAA ATATGCTATA TTAATACTTC	1560
AAAAATTACA CAAATAATAC ATAATCAGGT TTGAAAAATT TAAGACAACM SAARAAAAAA	1620
WYCMATCAC AMATATCCCA CACATTTTAT TATIMCTMCT MCWATTATTT TGWAGAGMCT	1680
GGGTCTCACY CYKTTGCTWA TGCTGGTCTT TGAACYCCYK GCCYCAARCA RTCCTSCITC	1740
ABCTCCCAA RGTGCTGGGG ATWATAGGCA TGARCTAACC GCACCCAGCC CCAGACATTT	1800
TAGTGTGTAA ATTCCTGGGC ATTTTTTCAA GGCATCATAC ATGTIAGCTG ACTGATGATG	1860
GTCAATTTAT TTTGTCCATG GTGTCAAGTT TCTCTTCAGG AGGAAAAGCA CAGAACTGGC	1920
CAACAATTGC TTGACTGTTC TTTACCATAC TGTTTAGCAG GAAACCAGTC TCAGTGTCCA	1980
ACTCTCTAAC CTTGGAAC TGAGAACTCT GAGGACAAA GAGCGGATAC AACCTCAAAA	2040
GACGTCTGTC TACATTGAAT TGGGTAAAGG TCTCAGGTTT TTTAAGTATT TAATAATAAT	2100
TGCTGGATT CTTATCTTAT AGTTTTGCCA AAAATCTTGG TCATAATTIG TATTTGTGGT	2160
AGGCAGCTTT GGGAAAGTGAA TTTTATGAGC CCTATGGTGA GTTATAAAAA ATGTAAAAGA	2220
CGCAGTTCCC ACCTTGAAGA ATCTTACTTT AAAAAGGGAG CAAAAGAGGC CAGGCATGGT	2280
GGCTCACACC TGTAATCCCA GCACTTTGGG AGGCCAAAAGT GGGTGGATCA CCTGAGGTCTG	2340
GGAGTTGAG ACCAGCCTAG CCAACATGGA GAAACTCTGT CTGTACCAA AAATAAAAAA	2400
TTAGCCAGGT GTGGTGGCAC ATAACGTGTA TCCCAGCTAC TCGGGAGGCT GAGGCAGGAG	2460
AATCACTTGA ACCCGGGAAG TGGAGGTTGC GGTGAACCGA GATCGCACCA TTGCACTCCA	2520
GCCTGGGCAA AAATAGCGAA ACTCCATCTA AAAAAAAAAA AGAGAGCAAA AGAAAGAMTM	2580
TCTGGTTTTA AMTMTGTGTA AATATGTTTT TGGAAAGATG GAGAGTAGCA ATAAGAAAAA	2640
ACATGATGGA TTGCTACAGT ATTTAGTTCC AAGATAAATT GTACTAGATG AGGAAGCCTT	2700
TTAAGAAGAG CTGAATTGCC AGGCGCAGTG GCTCACGCTT GTAATCCAG CACTTTGGGA	2760
GGCCGAGGTG GCGGATCAC CTGAGGTCGG GAGTTCAAGA CCAGCCTGAC CAACATGGAG	2820
AAACCCCATC TCTACTAAAA AAAAAAAAAA AAAAATTAGC CGGGGTGGTG GCTTATGCCT	2880
GTAATCCAG CTACTCAGGA GGCTGAGGCA GGAGAATCGC TTGAACCCAG GAAGCAGAGG	2940
TTGCAGTGAG CCAAGATCGC ACCATTGCAC TCCAGCCTAG GCAACAAGAG TGAAACTCCA	3000
TCTCAAAAAA AAAAAAAAAA AGCTGAATCT TGGCTGGGCA GGATGGCTCG TGCTGTAAAT	3060
CCTAACGCTT TGGAAAGACCG AGGCAGAAAG ATTGGTTGAG TCCACGAGTT TAAGACCAGC	3120
CTGGCCAACA TAGGGGAACC CTGTCTCTAT TTTTAAATA ATAATACATT TTTGGCCGGT	3180
GCGGTGGCTC ATGCCTGTAA TCCCAATACT TTGGGAGGCT GAGGCAGGTA GATCACCTGA	3240
GGTCAGAGTT CGAGACCAGC CTGGATAACC TGGTGAAACC CCTCTTTACT AAAAAATCAA	3300
AAAAAAAAAA AAATTAGCTG GGTGTGGTAG CACATGCTTG TAATCCCAGC TACTTGGGAG	3360
GCTGAGGCAG GAGAATCGCT TGAACCAAGG AGGCGGAGGT TACAATGAGC CAACACTACA	3420
CCACTGCACT CCAGCCTGGG CAATAGAGTG AGACTGCATC TCAAAAAAAT AATAATTTTT	3480
AAAAATAATA AATTTTTTTA AGCTTATAAA AAGAAAAGTT GAGGCCAGCA TAGTAGCTCA	3540
CATCTGTAAT CTCAGCAGTG GCAGAGGATT GCTTGAAGCC AGGAGTTTGA GACCAGCCTG	3600

109			5,753,441	110		
-continued						
GGCAACATAG	CAAGACCTCA	TCTCTACAAA	AAAATTTCTT	TTTTAAATTA	GCTGGGTGTG	3660
GTGGTGTGCA	TCTGTAGTCC	CAGCTACTCA	GGAGGCAGAG	GTGAGTGGAT	ACATTGAACC	3720
CAGGAGTTTG	AGGCTGTAGT	GAGCTATGAT	CAIGCCACTG	CACITCCAACC	TGGGTGACAG	3780
AGCAAGACCT	CAAAAAAAAA	AAAAAAAAAGA	GCTGCTGAGC	TCAGAAITCA	AACTGGGCTC	3840
TCAAATTGGA	TTTTCTTTTA	GAATATATTT	ATAATTAAAA	AGGATAGCCA	TCTTTTGAGC	3900
TCCCAGGCAC	CACCATCTAT	TTATCATAAC	ACTTACTGTT	TTCCCCCCTT	ATGATCATAA	3960
ATTCTAGAC	AACAGGCATT	GTAAAAATAG	TTATAGTAGT	TGATATTTAG	GAGCACITAA	4020
CTATATTCCA	GGCACTATTG	TGCTTTTCTT	GTATAACTCA	TTAGATGCTT	GTCAGACCTC	4080
TGAGATTGTT	CCTATTATAC	TTATTTTACA	GATGAGAAAA	TTAAGGCACA	GAGAAGTTAT	4140
GAAATTTTTT	CAAGGTATTA	AACCTAGTAA	GTGGCTGAGC	CATGATTCAA	ACCTAGGAAG	4200
TTAGATGTCA	GAGCCTGTGC	TTTTTTTTTG	TTTTTGTTTT	TGTTTTCAGT	AGAAACGGGG	4260
GTCTCACTTT	GTTGGCCAGG	CTGGTCTTGA	ACTCCTAACC	TCAAATAATC	CACCCATCTC	4320
GGCCTCCTCA	AGTGCTGGGA	TTACAGGTGA	GAGCCACTGT	GCCTGGCGAA	GCCCATGCCT	4380
TTAACCACIT	CTCTGTATTA	CATACTAGCT	TAAGTAGCAT	TGTACCTGCC	ACAGTAGATG	4440
CTCAGTAAAT	ATTCTAGTGT	GAATATCTGT	TTTTCAACAA	GTACATTTTT	TTAACCCTTT	4500
TAATTAAGAA	AACITTTTAT	GATTTATTTT	TTGGGGGGAA	ATTTTTTAGG	ATCTGATTCT	4560
TCTGAAGATA	CCGTTAATAA	GGCAACTTAT	TGCAGGTGAG	TCAAAGAGAA	CCTTTGTCTA	4620
TGAAGCTGGT	ATTTTCCTAT	TTAGTTAATA	TTAAGGATTG	ATGTTTCTCT	CTTTTTAAAA	4680
ATATTTTAAAC	TTTTATTTTA	GGTTCAGGGA	TGTATGTGCA	GTTTGTTATA	TAGGTAAACA	4740
CACGACTTGG	GATTTGGTGT	ATAGATTTTT	TTATCATCCT	GGGTACTAAG	CATACCCAC	4800
AGTTTTTTGT	TTGCTTTCTT	TCTGAATTTT	TCCCTCTTCC	CACCTTCCTC	CCTCAAGTAG	4860
GCTGGTGTIT	CTCCAGACTA	GAATCATGGT	ATTGGAAGAA	ACCTTAGAGA	TCATCTAGTT	4920
TAGTTCTCTC	ATTTTATAGT	GGAGGAAATA	CCCTTTTTGT	TTGTTGGATT	TAGTTATTAG	4980
CACTGTCCAA	AGGAATTTAG	GATAACAGTA	GAAGCTGTGA	CATGCTTGCT	TCTAGCAGAT	5040
TGTTCTCTAA	GTTCCTCATA	TACAGTAATA	TTGACACAGC	AGTAATTGTG	ACTGATGAAA	5100
ATGTTCAAGG	ACTTCATTTT	CAACTCTTTC	TTTCCTCTGT	TCCTTATTTT	CACATATCTC	5160
TCAAGCTTTG	TCTGTATGTT	ATATAATAAA	CTACAAGCAA	CCCCAACTAT	GTTACCTACC	5220
TTCTTTAGGA	ATTATTGCTT	GACCCAGGTT	TTTTTTTTTT	TTTTTTTGGG	GACGGGGTCT	5280
TGCCCTGTTG	CCAGGATGGA	GTGTAGTGGC	GCCATCTCGG	CTCACTGCAA	TCTCCAACCTC	5340
CCTGGTTCAA	GCGATTCTCC	TGCTCTCAATC	TCACGAGTAG	CTGGGACTAC	AGGTATACAC	5400
CACCACGCCC	GGTTAATTGA	CCATTCCATT	TCTTTCITTC	TCTCTTTTTT	TTTTTTTTTT	5460
TTGAGACAGA	GTCTTGCTCT	GTTGCCCAGG	CTGGAGTACA	GAGGTGTGAT	CTCACCTCTC	5520
CGCAACGTCT	GCCTCCCAGG	TTGAAGCCAT	ACTCCTGCCT	CAGCCTCTCT	AGTAGCTGGG	5580
ACTACAGGCG	CGCGCCACCA	CACCCGGCTA	ATTTTTGTAT	TTTTAGTAGA	GATGGGGTTT	5640
CACCATGTTG	GCCAGGCTGG	TCTTGAATCT	ATGACCTCAA	GTGGTCCACC	CGCCTCAGCC	5700
TCCCAAAGTG	CTGGAATTAC	AGGCTTGAGC	CACCGTGCCC	AGCAACCATT	TCATTTCAAC	5760
TAGAAGTTTC	TAAAGGAGAG	AGCAGCTTTC	ACTAACTAAA	TAAGATTGGT	CAGCTTTCTG	5820
TAATCGAAAG	AGCTAAAATG	TTTGATCTTG	GTCAATTTGAC	AGTTCTGCAT	ACATGTAAC	5880
AGTGTTTCTT	ATTAGGACTC	TGTCTTTTCC	CTATAGTGTG	GGAGATCAAG	AATTGTTACA	5940
AATCACCCCT	CAAGGAACCA	GGGATGAAAT	CAGTTTGGAT	TCTGCAAAAA	AGGGTAATGG	6000

5,753,441

111

112

-continued

C A A A G T T T G C	C A A C T T A A C A	G G C A C T G A A A	A G A G A G T G G G	T A G A T A C A G T	A C T G T A A T T A	6 0 6 0
G A T T A T T C T G	A A G A C C A T T T	G G G A C C T T T A	C A A C C C A C A A	A A T C T C T T G G	C A G A G T T A G A	6 1 2 0
G T A T C A T T C T	C T G T C A A A T G	T C G T G G T A T G	G T C T G A T A G A	T T T A A A T G G T	A C T A G A C T A A	6 1 8 0
T G T A C C I A T A	A T A A G A C C T T	C I T G T A A C T G	A T T G T T G C C C	T T T C G C T T T T	T T T T T T G T T T	6 2 4 0
G T T T G T T T G T	T T T T T T T T G A	G A T G G G G T C T	C A C T C T G T T G	C C C A G G C T G G	A G T G C A G T G A	6 3 0 0
T G C A A T C T T G	G C T C A C T G C A	A C C T C C A C C T	C C A A A G G C T C	A A G C T A T C C T	C C C A C T T C A G	6 3 6 0
C C T C C T G A G T	A G C T G G G A C T	A C A G G C G C A T	G C C A C C A C A C	C C G G T T A A T T	T T T T G T G G T T	6 4 2 0
T T A T A G A G A T	G G G G T T T C A C	C A T G T T A C C G	A G G C T G G T C T	C A A A C T C C T G	G A C T C A A G C A	6 4 8 0
G T C T G C C C A C	T T C A G C C T C C	C A A A G T G C T G	C A G T T A C A G G	C T T G A G C C A C	T G T G C C T G G C	6 5 4 0
C T G C C C T T T A	C T T T T A A T T G	G T G T A T T T G T	G T T T C A T C T T	T T A C C T A C T G	G T T T T T A A A T	6 6 0 0
A T A G G G A G T G	G T A A G T C T G T	A G A T A G A A C A	G A G T A T T A A G	T A G A C T T A A T	G G C C A G T A A T	6 6 6 0
C T T T A G A G T A	C A T C A G A A C C	A G T T T T C T G A	T G G C C A A T C T	G C T T T T A A T T	C A C T C T T A G A	6 7 2 0
C G T T A G A G A A	A T A G G T G T G G	T T T C T G C A T A	G G G A A A A T T C	T G A A A T T A A		6 7 8 0

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 4249 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(i v) ANTI-SENSE: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:21:

GATCCTAAGT	GGAAATAATC	TAGGTAATA	GGAATTAAT	GAAAGAGTAT	GAGCTACATC	60
TTCAGTATAC	TGGTAGTTT	ATGAGGTTAG	TTTCTCTAAT	ATAGCCAGTT	GGTTGATTTT	120
CACCTCCAAG	GTGTATGAAG	TATGTATTTT	TTTAATGACA	ATTCA GTTTT	TGAGTACCTT	180
GTTATTTTTT	TATATTTTCA	GCTGCTTGTG	AATTTTCTGA	GACGGATGTA	ACAAATACTG	240
AACATCATCA	ACCCAGTAAT	AATGATTTGA	ACACCACTGA	GAAGCGTGCA	GCTGAGAGGC	300
ATCCAGAAAA	GTATCAGGGT	AGTTCTGTTT	CAAACCTTGA	TGTGGAGCCA	TGTGGCACAA	360
ATACTCATGC	CAGCTCATT	CAGCATGAGA	ACAGCAGTTT	ATTACTCACT	AAAGACAGAA	420
TGAATGTAGA	AAAGGCTGAA	TTCTGTAATA	AAAGCAAACA	GCCTGGCTTA	GCAAGGAGCC	480
AACATAACAG	ATGGGCTGGA	AGTAAGGAAA	CATGTAATGA	TAGGCGGACT	CCCAGCACAG	540
AAAAAAAGGT	AGATCTGAAT	GCTGATCCCC	TGTGTGAGAG	AAAAGAATGG	AATAAGCAGA	600
AACTGCCATG	CTCAGAGAAT	CCTAGAGATA	CTGAAGATGT	TCCTTGGATA	ACACTAAATA	660
GCAGCATTCA	GAAAGTTAAT	GAGTGGTTTT	CCAGAAGTGA	TGAAC TGTTA	GGTTCTGATG	720
ACTCACATGA	TGGGGAGTCT	GAATCAAATG	CCAAAGTAGC	TGATGTATTG	GACGTTCTAA	780
ATGAGGTAGA	TGAATATTCT	GGTTCTTCAG	AGAAAAATAGA	CTTACTGGCC	AGTGATCCTC	840
ATGAGGCTTT	AATATGTAAA	AGTGAAAGAG	TTCATCCAA	ATCAGTAGAG	AGTAATATTG	900
AAGGCCAAAT	ATTTGGGAAA	ACCTATCGGA	AGAAGGCAAG	CCTCCCCAAC	TTAAGCCATG	960
TAAC TGAAAA	TCTAATTATA	GGAGCATTTG	TTACTGAGCC	ACAGATAATA	CAAGAGCGTC	1020

113			5,753,441		114		
						-continued	
CCCTCACAAA	TAAATTAAAG	CGTAAAAGGA	GACCTACATC	AGGCCTTCAT	CCTGAGGATT	1080	
TTATCAAGAA	AGCAGATTTG	GCAGTTCAAA	AGACTCCTGA	AATGATAAAT	CAGGGAAC TA	1140	
ACCAAACGGA	GCAGAATGGT	CAAGTGATGA	ATATTACTAA	TAGTGGTCAT	GAGAATAAAA	1200	
CAAAAGGTGA	TTCTATTGAG	AATGAGAAAA	ATCCTAACCC	AATAGAATCA	CTCGAAAAAG	1260	
AATCTGCTTT	CAAAACGAAA	GCTGAACCTA	TAAGCAGCAG	TATAAGCAAT	ATGGAAC TCG	1320	
AATTAATAT	CCACAATTCA	AAAGCACCTA	AAAAGAATAG	GCTGAGGAGG	AAGTCTTCTA	1380	
CCAGGCATAT	TCATGCGCTT	GAAGTAGTAG	TCAGTAGAAA	TCTAAGCCCA	CCTAATTGTA	1440	
CTGAATTGCA	AATTGATAGT	TGTTCTAGCA	GTGAAGAGAT	AAAGAAAAAA	AAGTACAACC	1500	
AAATGCCAGT	CAGGCACAGC	AGAAACCTAC	AACTCATGGA	AGGTAAAGAA	CCTGCAACTG	1560	
GAGCCAAGAA	GAGTAACAAG	CCAAATGAAC	AGACAAGTAA	AAGACATGAC	AGCGATACTT	1620	
TCCCAGAGCT	GAAGTTAACA	AATGCACCTG	GTTCTTTTAC	TAAGTGTTCA	AATACCACTG	1680	
AACTTAAAGA	ATTTGTCAAT	CCTAGCCTTC	CAAGAGAAGA	AAAAGAAGAG	AACTAGAAAC	1740	
AGTTAAAGTG	TCTAATAATG	CTGAAGACCC	CAAAGATCTC	ATGTTAAGTG	GAGAAAGGGT	1800	
TTTGCAAACT	GAAAGATCTG	TAGAGAGTAG	CAGTATTTCA	TTGGTACCTG	GIAC TGATTA	1860	
TGGCACTCAG	GAAAGTATCT	CGTTACTGGA	AGTTAGCACT	CTAGGGAAGG	CAAAAACAGA	1920	
ACCAAATAAA	TGTGTGAGTC	AGTGTGCAGC	ATTTGAAAAAC	CCCAAGGGAC	TAATTCATGG	1980	
TTGTTCCAAA	GATAATAGAA	ATGACACAGA	AGGCCTTAAG	TATCCATTGG	GACATGAAGT	2040	
TAACCACAGT	CGGGAAACAA	GCATAGAAAT	GGAAGAAAAGT	GAAC TTGATG	CTCAGTATTT	2100	
GCAGAATACA	TTCAAGGTTT	CAAAGCGCCA	GTCAATTTGCT	CCGTTTTTCAA	ATCCAGGAAA	2160	
TGCAGAAGAG	GAATGTGCAA	CATTCTCTGC	CCACTCTGGG	TCCTTAAAGA	AACAAAAGTCC	2220	
AAAAGTCACT	TTTGAATGTG	AACAAAAGGA	AGAAAATCAA	GGAAAGAATG	AGTCTAATAT	2280	
CAAGCCTGTA	CAGACAGTTA	ATATCACTGC	AGGCCTTCCT	GTGGTTGGTC	AGAAAGATAA	2340	
GCCAGTTGAT	AATGCCAAAT	GTAGTATCAA	AGGAGGCTCT	AGGTTTTGTC	TATCATCTCA	2400	
GTTTCAGAGGC	AACGAAACTG	GACTCATTAC	TCCAAATAAA	CATGGACTTT	TACAAAACCC	2460	
ATATCGTATA	CCACCACTTT	TTCCCATCAA	GTCAATTTGT	AAAACTAAAT	GTAAGAAAAA	2520	
TCTGCTAGAG	GAAAACTTTG	AGGAACATTG	AATGTCACCT	GAAAGAGAAA	TGGGAAATGA	2580	
GAACATTCCA	AGTACAGTGA	GCACAATTAG	CCGTAATAAC	ATTAGAGAAA	ATGTTTTTAA	2640	
AGAAGCCAGC	TCAAGCAATA	TTAATGAAGT	AGGTTCCAGT	ACTAATGAAG	TGGGCTCCAG	2700	
TATTAATGAA	ATAGGTTCCA	GTGATGAAAA	CATTCAAGCA	GAAGTAGGTA	GAAACAGAGG	2760	
GCCAAAATTG	AATGCTATGC	TTAGATTAGG	GGTTTTGCAA	CCTGAGGTCT	ATAAACAAAG	2820	
TCTTCTGGA	AGTAATTGTA	AGCATCCTGA	AATAAAAAAG	CAAGAATATG	AAGAAGTAGT	2880	
TCAGACTGTT	AATACAGATT	TCTCTCCATA	TCTGATTTCA	GATAACTTAG	AACAGCCTAT	2940	
GGGAAGTAGT	CATGCATCTC	AGGTTTGTTC	TGAGACACCT	GATGACCTGT	TAGATGATGG	3000	
TGAAATAAAG	GAAGATACTA	GTTTTGCTGA	AAATGACATT	AAGGAAAGTT	CTGCTGTTTT	3060	
TAGCAAAAAGC	GTCCAGAAAAG	GAGAGCTTAG	CAGGAGTCCT	AGCCCTTTCA	CCCATACACA	3120	
TTTGGCTCAG	GGTTACCGAA	GAGGGGCCAA	GAAATTAGAG	TCCTCAGAAG	AGAACTTATC	3180	
TAGTGAGGAT	GAAGAGCTTC	CCTGCTTCCA	ACACTTGTTA	TTTGGTAAAG	TAAACAATAT	3240	
ACCTTCTCAG	TCTACTAGGC	ATAGCACCGT	TGCTACCGAG	TGTCTGTCTA	AGAACACAGA	3300	
GGAGAATTTA	TTATCATTGA	AGAATAGCTT	AAATGACTGC	AGTAACCAGG	TAATATTGGC	3360	
AAAGGCATCT	CAGGAACATC	ACCTTAGTGA	GGAAACAAAA	TGTTCTGCTA	GCTTGTTTTT	3420	

5,753,441

115

116

-continued

TTACACAGTGC	AGTGAATTGG	AAGACTTGAC	TGCAAAATACA	AACACCCAGG	ATCCTTTCTT	3480
GATTGGTTCT	TCCAAACAAA	TGAGGCATCA	GTC TGAAAGC	CAGGGAGTTG	GCTGAGTGA	3540
CAAGGAATTG	GTTTCAGATG	ATGAAGAAAG	AGGAACGGGC	TTGGAAGAAA	ATAATCAAGA	3600
AGAGCAAAGC	ATGGATTCAA	ACTTAGGTAT	TGGAACCAGG	TTTTTGTTT	TGCCCCAGTC	3660
TATTTATAGA	AGTGAGCTAA	ATGTTTATGC	TTTTGGGGAG	CACATTTTAC	AAATTTCCAA	3720
GTATAGTTAA	AGGAACTGCT	TCITAAACTT	GAAACATGTT	CCCTCTAAGG	TGCTTTTCAT	3780
AGAAAAAAGT	CCTTCACACA	GCTAGGACGT	CATCTTTGAC	TGAATGAGCT	TTAACATCCT	3840
AATTACTGGT	GGACTTACTT	CTGGTTTCAT	TTTATAAAGC	AAATCCCGGT	GTCCCAAAGC	3900
AAGGAATTTA	ATCATTTTGT	GTGACATGAA	AGTAAATCCA	GTCTTGCCAA	TGAGAAGAAA	3960
AAGACACAGC	AAGTTGCAGC	GTTTATAGTC	TGCTTTTACA	TCIGAACCTC	TGTTTTTGTT	4020
ATTTAAGGTG	AAGCAGCATC	TGGGTGTGAG	AGTGAAACAA	GCGTCTCTGA	AGACTGCTCA	4080
GGGCTATCCT	CTCAGAGTGA	CATTTTAAAC	ACTCAGGTAA	AAAGCGTGTG	TGTGTGTGCA	4140
CATGCGTGTG	TGTGGTGTCC	TTTGCAATTCA	GTAGTATGTA	TCCCACATTC	TTAGGTTTGC	4200
TGACATCATC	TCTTTGAATT	AATGGCACAA	TIGTTTGTGG	TTCATTGTC		4240

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 710 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(i v) ANTI-SENSE: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: *Homo sapiens*

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:22:

NONGAATGTA	ATCCTAATAT	TTCNCNCCNA	CTTAAAAGAA	TACCACTCCA	ANGGCATCNC	60
AATACATCAA	TCAATTGGGG	AATTGGGATT	TTCCTCNCCT	AACATCANTG	GAATAATTTT	120
ATGGCATTAA	TTGCATGAAT	GTGGTTAGAT	TAAAAGGTGT	TCATGCTAGA	ACTTGTAGTT	180
CCATACTAGG	TGATTTCAAT	TCTGTGCTA	AAATTAATTT	GTATGATATA	TTNTCATTTA	240
ATGGAAAGCT	TCTCAAAGTA	TTTCATTTTC	TTGGTACCAT	TTATCGTTTT	TGAAGCAGAG	300
GGATACCATG	CAACATAACC	TGATAAAGCT	CCAGCAGGAA	ATGGCTGAAC	TAGAAGCTGT	360
GTTAGAACAG	CATGGGAGCC	AGCCTTCTAA	CAGCTACCCCT	TCCATCATAA	GTGACTCTTC	420
TGCCCTTGAG	GACCTGCGAA	ATCCAGAACA	AAGCACATCA	GAAAAAGGTG	TGTATTGTTG	480
GCCAAACACT	GATATCTTAA	GCAAAATTCT	TTCCTTCCCC	TTTATCTCCT	TCTGAAGAGT	540
AAGGACCTAG	CTCCAACATT	TTATGATCCT	TGCTCAGCAC	ATGGGTAATT	ATGGAGCCTT	600
GGTTCTTGTC	CCTGCTCACA	ACTAATATAC	CAGTCAGAGG	GACCCAAGGC	AGTCATTTCAT	660
GTTGTCACT	GAGATACCTA	CAACAAGTAG	ATGCTATGGG	GAGCCCCATGG		710

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 473 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double

A000328

5,753,441

117

118

-continued

(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(i v) ANTI-SENSE: NO

(y i) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:23:

CCATTGGTGC	TAGCATCTGT	CTGTTGCATT	GCTTIGITTT	ATAAAATTCT	GCCTGATATA	60
CTTGTTAAAA	ACCAATTTGT	GTATCATAGA	TTGATGCTTT	TGAAAAAAT	CAGTATTCTA	120
ACCTGAATTA	TCACTATCAG	AACAAAGCAG	TAAAGTAGAT	TTGTTTTCTC	ATTCATTTA	180
AAGCAGTATT	AAC TTCACAG	AAAAGTAGTG	AATACCCTAT	AAGCCAGAAT	CCAGAAGGCC	240
TTTCTGCTGA	CAAGTTTGAG	GTGTCTGCAG	ATAGTTCTAC	CAGTAAAAAT	AAAGAACCAG	300
GAGTGGAAAG	GTAAGAAACA	TCAATGTAAA	GATGCTGTGG	TATCTGACAT	CTTTATTTAT	360
ATTGAACTCT	GATTGTTAAT	TTTTTTCACC	ATACTTTCTC	CAGTTTTTTT	GCATACAGGC	420
ATTTATACAC	TTTTATTGCT	CTAGGATACT	TCTTTIGITT	AATCCTATAT	AGG	473

(2) INFORMATION FOR SEO ID NO:24:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 421 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(i v) ANTI-SENSE: NO

(v i) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:24:

GGATAAGNTC	AAGAAGATATT	TTGATAGGTG	ATGCAGTGAT	NAATTGNGAA	AATTINCTGC	60
CTGCTTTTAA	TCTTCCCCCG	TTCTTTCTTC	CTNCCITCCCT	CCCTTCCTNC	CTCCCGTCCT	120
TNCCTTTCCT	TTCCTCCCT	TCCNCCTTCT	TTCNCTCTNT	CTTTCCTTTC	TTTCCTGTCT	180
ACCTTTCCTT	CCTTCCTCCC	TTCTTTTCT	TTTCTTTCCT	TCCTTTCCTT	TTCTTTCCTT	240
TCITTCCITT	CCTTTCCTTC	TTGACAGAGT	CTTGCTCTGT	CACTCAGGCT	GGAGTGAGT	300
GGCGTGATCT	CGNCTCACTG	CAACCTCTGT	CTCCCAGGTT	CAAGCAATTT	TCCTGCCTCA	360
GCCTCCCGAG	TAGCTGAGAT	TACAGGCGCC	AGCCACCACA	CCCAGCTACT	GACCTGCTTT	420
T						421

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 997 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(i v) ANTI-SENSE: NO

5,753,441

119

120

-continued

(v i) ORIGINAL SOURCE:

(A) ORGANISM: *Homo sapiens*

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:25:

ATAACAGCTGG	GAGATATGGT	GCCTCAGACC	AACCCCATGT	TATATGTCAA	CCCTGACATA	60
TGGCAGGCA	ACATGAATCC	AGACTTCTAG	GCTGTCAATG	GGGCTCTTTT	TGCCAGTCA	120
TTTCTGATCT	CCTGACATG	AGCTGTTTCA	TTTATGCTTT	GGCTGCCCAG	CAAATATGAT	180
TGTGCTTTT	ACAATGGTG	GCGATGGTTT	TCCTCTTCCA	TTTATCTTTC	TAGGTTCATCC	240
CCTTCTAAAT	GCCCATCATT	AGATGATAGG	TGGTACATGC	ACAGTTGCTC	TGGGAGTCTT	300
CAGAATAGAA	ACTACCCATC	TCAAGAGGAG	CTCATTAAGG	TGTTGATGT	GGAGGAGCAA	360
CAGCTGGAAG	AGTCTGGGCC	ACACGATTTG	ACGGAAACAT	CCTACTTGCC	AAGGCAAGAT	420
CTAGGTAATA	TTTCATCTGC	TGTATTGGAA	CAAACTCTYT	GATTTTACTC	TGAATCCTAC	480
ATAAAGATAT	TCGGTTAAC	CAACTTTTAG	ATGTACTAGT	CTATCATGGA	CACTTTTGTT	540
ATACTTAATT	AAGCCCACTT	TAGAAAAATA	GCTCAAGTGT	TAATCAAGGT	TACTTGAAA	600
ATTATTGAAA	CTGTTAATCC	ATCTATATTT	TAATTAATGG	TTAACTAAT	GATTTTGAGG	660
ATGWGGGAGT	CKTGGTGTAC	TCTAMATGTA	TTATTTCAGG	CCAGGCATAG	TGGCTCACGC	720
CTGGTAATCC	CAGTAYYCMR	GAGCCCCAGG	CAGGTGGAGC	CAGCTGAGGT	CAGGAGTTCA	780
AGACCTGTCT	TGGCCAACAT	GGGNGAAACC	CTGCTTCTT	CTTAAAAAAN	ACAAAAAAA	840
TTAACTGGGT	TGTGCTTAGG	TGNATGCCCC	GNATCCTAGT	TNTTCTTGNG	GGTTGAGGGA	900
GGAGATCACN	TGGACCCCG	GAGGGGNGGG	TGGGGGNGAG	CAGGNCAAAA	CACNGACCCA	960
GCTGGGGTGG	AAGGGAAGCC	CATCTNAAAA	AANNITN			990

(2) INFORMATION FOR SEO ID NO:26:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 639 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(i v) ANTI-SENSE: NO

(v i) ORIGINAL SOURCE:

(A) ORGANISM: *Homo sapiens*

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:26:

TTTTTAGGAA	ACAAGCTACT	TTGGATTTC	ACCAACACCT	GTATTCATGT	ACCCATTTTT	60
CTCTTAACCT	AACCTTATIG	GTCTTTTTAA	TCTTAACAG	AGACCAGAAC	TTGTAAATTC	120
AACATTCATC	GTGTGTAAA	TTAAACTTCT	CCCATTCCIT	TCAGAGGGAA	CCCCTTACCT	180
GGAATCTGGA	ATCAGCCTCT	TCTCTGATGA	CCCTGAATCT	GATCCTTCTG	AAGACAGAGC	240
CCCAGAGTCA	GCTCGTGTIG	GCAACATACC	ATCTTCAACC	TCTGCATTGA	AAGTTCCCCA	300
ATTGAAAGTT	GCAGAATCTG	CCCAGAGTCC	AGCTGCTGCT	CATACTACTG	ATACTGCTGG	360
GTATAATGCA	ATGGAAGAAA	GTGTGAGCAG	GGAGAAGCCA	GAATTGACAG	CTTCAACAGA	420
AAGGGTCAAC	AAAAGAATGT	CCATGGTGGT	GTCTGGCCTG	ACCCCAGAAG	AATTTGTGAG	480
TGTATCCAAT	TGTATCTCCC	TAATGACTAA	GACTTAACAA	CATTCTGGAA	AGAGTTTTAT	540
GTAGGTATTG	TCAATTAATA	ACCTAGAGGA	AGAAATCTAG	AAAACAATCA	CAGTTCTGTG	600
TAATTTAATT	TCGATTACTA	ATTTCTGAAA	ATTTAGAA			630

5,753,441

121

122

-continued

(2) INFORMATION FOR SEQ ID NO:27:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 922 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(i v) ANTI-SENSE: NO

(v i) ORIGINAL SOURCE:

- (A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:27:

NCCCNCCCC	CNAATCTGAA	ATGGGGGTAA	CCCCCCCCCA	ACCGANACNT	GGGTNGCNTA	60
GAGANTTTAA	TGGCCCNTTC	TGAGGNACAN	AAGCTTAAGC	CAGGNGACGT	GGANCNATGN	120
GTGTGTTNTT	GTITGGTTAC	CTCCAGCCTG	GGTGACAGAG	CAAGACTCTG	TCTAAAAAAA	180
AAAAAAAAAA	AAATCGACTT	TAAATAGTTC	CAGGACACGT	GTAGAACGTG	CAGGATTGCT	240
ACGTAGGTAA	ACATATGCCA	TGGTGGGATA	ACTAGTATTC	TGAGCTGTGT	GCTAGAGGTA	300
ACTCATGATA	ATGGAATATT	TGATTTAATT	TCAGATGCTC	GTGTACAAAGT	TTGCCAGAAA	360
ACACCAATC	ACTTTAACTA	ATCTAATTAC	TGAAGAGACT	ACTCATGTTG	TTATGAAAAC	420
AGGTATACCA	AGAACCCTTT	CAGAATACCT	TGCATCTGCT	GCATAAAACC	ACATGAGGCG	480
AGGCACGGTG	GCGCATGCCCT	GTAATCGCAG	CACCTTGGGA	GGCCGAGGCG	GGCAGATCAC	540
GAGATTAGGA	GATCGAGACC	ATCCTGGCCA	GCATGGTGAA	ACCCCGTCTC	TACTANNAAA	600
TGGNAAAATT	ANCTGGGTGT	GGTCGCGTGC	NCCTGTAGTC	CCAGCTACTC	GTGAGGCTGA	660
GGCAGGAGAA	TCACCTGAAC	CGGGGAAATG	GAGGTTTCAG	TGAGCAGAGA	TCATNCCCTT	720
NCATTCCAGC	CTGGCGACAG	AGCAAGGCTC	CGTCNCCNAA	AAAATAAAAA	AAAACGTGAA	780
CAAATAAGAA	TATTTGTTGA	GCATAGCATG	GATGATAGTC	TTCTAATAGT	CAATCAATTA	840
CTTTATGAAA	GACAAATAAT	AGTTTTTGCTG	CTTCCTTACC	TCCTTTTGTT	TTGGGTTAAG	900
ATTTGGAGTG	TGGGCCAGGC	AC				922

(2) INFORMATION FOR SEQ ID NO:28:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 867 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(i v) ANTI-SENSE: NO

(v i) ORIGINAL SOURCE:

- (A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:28:

GATCTATAGC	TAGCCTTGCC	GTCTAGAAGA	TGGGTGTTGA	GAAGAGGGAG	TGGAAAGATA	60
TTTCCTCTGG	TCCTAACTTC	ATATCAGCCT	CCCCTAGACT	TCCAAATATC	CATACCTGCT	120
GGTTATAATT	AGTGGTGTTT	TCAGCCTCTG	ATTCTGTAC	CAGGGGTTTT	AGAATCATAA	180
ATCCAGATTG	ATCTTGGGAG	TGTAAAAAAC	TGAGGCTCTT	TAGCTTCTTA	GGACAGCACT	240

5,753,441

123

124

-continued

T C C T G A T T T T	G T T T T C A A C T	T C T A A T C C T T	T G A G T G T T T T	T C A T T C T G C A	G A T G C T G A G T	3 0 0
T T G T G T G T G A	A C G G A C A C T G	A A A T A T T T T C	T A G G A A T T G C	G G G A G G A A A A	T G G G T A G T T A	3 6 0
G C T A T T T C T G	T A A G T A T A A T	A C T A T T T C T C	C C C T C C T C C C	T T T A A C A C C T	C A G A A T T G C A	4 2 0
T T T T T A C A C C	T A A C A T T T A A	C A C C T A A G G T	T T T T G C T G A T	G C T G A G T C T G	A G T T A C C A A A	4 8 0
A G G T C T T T A A	A T T G T A A T A C	T A A A C T A C T T	T T A T C T T T A A	T A T C A C T T T G	T T C A A G A T A A	5 4 0
G C T G G T G A T G	C T G G G A A A A T	G G G T C T C T T T	T A T A A C T A A T	A G G A C C T A A T	C T G C T C C T A G	6 0 0
C A A T G T T A G C	A T A T G A G C T A	G G G A T T T A T T	T A A T A G T C G G	C A G G A A T C C A	T G T G C A R C A G	6 6 0
N C A A A C T T A T	A A T G T T T A A A	T T A A A C A T C A	A C T C T G T C T C	C A G A A G G A A A	C T G C T G C T A C	7 2 0
A A G C C T T A T T	A A A G G G C T G T	G G C T T T A G A G	G G A A G G A C C T	C T C C T C T G T C	A T T C T T C C T G	7 8 0
T G C T C T T T T G	T G A A T C G C T G	A C C T C T C T A T	C T C C G T G A A A	A G A G C A C G T T	C T T C T G C T G T	8 4 0
A T G T A A C C T G	T C T T T T C T A T	G A T C T C T				8 6 7

(2) INFORMATION FOR SEQ ID NO:29:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 561 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(i v) ANTI-SENSE: NO

(v i) ORIGINAL SOURCE:

- (A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:29:

N A A A A A C G G G	G N N G G G A N T G	G G C C T T A A A N	C C A A A G G G C N	A A C T C C C C A A	C C A T T N A A A A	6 0
A N T G A C N G G G	G A T T A T T A A A	A N C G G C G G G A	A A C A T T T C A C	N G C C C A A C T A	A T A T T G T T A A	1 2 0
A T T A A A A C C A	C C A C C N C T G C	N C C A A G G A G G	G A A A C T G C T G	C T A C A A G C C T	T A T T A A A G G G	1 8 0
C T G T G G C T T T	A G A G G G A A G G	A C C T C T C C T C	T G T C A T T C T T	C C T G T G C T C T	T T T G T G A A T C	2 4 0
G C T G A C C T C T	C T A T G T C C G T	G A A A A G A G C A	C G T T C T I C G T	C T G T A T G T A A	C C T G T C T T T T	3 0 0
C T A T G A T C T C	T T T A G G G G T G	A C C C A G T C T A	T T A A A G A A A G	A A A A A T G C T G	A A T G A G G T A A	3 6 0
G T A C T T G A T G	T T A C A A A C T A	A C C A G A G A T A	T T C A T T C A G T	C A T A T A G T T A	A A A A T G T A T T	4 2 0
T G C T T C C T T C	C A T C A A T G C A	C C A C T T T C C T	T A A C A A T G C A	C A A A T T T T C C	A T G A T A A T G A	4 8 0
G G A T C A T C A A	G A A T T A T G C A	G G C C T G C A C T	G T G G G T C A T A	C C T A T A A T C C	C A G C G C T T T G	5 4 0
G G A G G C T G A G	G C G C T T G G A T	C				5 6 1

(2) INFORMATION FOR SEQ ID NO:30:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 567 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(i v) ANTI-SENSE: NO

(v i) ORIGINAL SOURCE:

- (A) ORGANISM: Homo sapiens

5,753,441

125

126

-continued

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:30:

AATTTTITTTGT	ATTTTITAGTA	GAGATGAGGT	TCACCATGTT	GOTCTAGATC	TGGTGTGAA	60
CGTCCTGACC	TCAAGTGATC	TGCCAGCCTC	AGTCTCCCAA	AGTGCTAGGA	TTACAGGGGT	120
GAGCCACTGC	GCCTGGCCTG	AATGCCTAAA	ATATGACGTG	TCTGCTCCAC	TTCCATTGAA	180
GGAAGCTTCT	CTTTCICTTA	TCCTGATGGG	TTGTGTTTGG	TTTCTTTCAG	CATGATTTTG	240
AAGTCAGAGG	AGATGTGGTC	AATGGAAGAA	ACCACCAAGG	TCCAAAGCGA	GCAAGAGAAT	300
CCCAGGACAG	AAAGGTAAAG	CTCCCTCCCT	CAAGTTGACA	AAAATCTCAC	CCCACCACTC	360
TGTATTCCAC	TCCCCTTTGC	AGAGATGGGC	CGCTTCATTT	TGTAAGACTT	ATTACATACA	420
TACACAGTGC	TAGATACTTT	CACACAGGTT	CTTTTTTCAC	TCTTCCATCC	CAACCACATA	480
AATAAGTATT	GTCTCTACTT	TATGAATGAT	AAAATAAGA	GATTTAGAGA	GGCTGTGTAA	540
TTTGGATTCC	CGICTCGGGT	TCAGATC				567

(2) INFORMATION FOR SEO ID NO:31:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 633 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(i v) ANTI-SENSE: NO

(v i) ORIGINAL SOURCE:

(A) ORGANISM: *Homo sapiens*

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:31:

TTGGCCTGAT	TGGTGACAAA	AGTGAGATGC	TCAGTCCTTG	AATGACAAAG	AATGCCTGTA	60
GAGTTGCAGG	TCCAAC TACA	TATGCAC TTC	AAGAAGATCT	TCTGAAATCT	AGTAGTGTTC	120
TGGACATTGG	ACTGCTTGTC	CCTGGGAAGT	AGCAGCAGAA	ATGATCGGTG	GTGAACAGAA	180
GAAAAAGAAA	AGCTCTTCCT	TTTGTAAAGT	CTGTTTTTTG	AATAAAAGCC	AATATTCTTT	240
TATAACTAGA	TTTTCTTTCT	CTCCATTCCC	CTGTCCCTCT	CTCTTCTCT	CTTCTTCCAG	300
ATCTTCAGGG	GGCTAGAAAT	CTGTTGCTAT	GGGCCCTTCA	CCAACATGCC	CACAGGTAAG	360
AGCCTGGGAG	AACCCAGAG	TTCCAGCACC	AGCCTTTGTC	TTACATAGTG	GAGTATTATA	420
AGCAAGGTCC	CACGATGGGG	GTTCTCTAGA	TTGCTGAAAT	GTTCTAGAGG	CTATTCTATT	480
TCTCTACCAC	TCTCAAACA	AAACAGCACC	TAAATGTTAT	CCTATGGCAA	AAAAAACTA	540
TACCTTGTC	CCCTTCTCAA	GAGCATGAAG	GTGGTTAATA	GTTAGGATT	AGTATGTTAT	600
GTGTTTCAGAT	GGCGTTGAGC	TGCTGTTAGT	GCC			633

(2) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 470 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(i v) ANTI-SENSE: NO

5,753,441

127

128

-continued

(v i) ORIGINAL SOURCE:

(A) ORGANISM: *Homo sapiens*

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:32:

TTTGAGAGAC	TATCAAACCT	TATACCAAGT	GGCCTTATGG	AGACTGATAA	CCAGAGTACA	60
TGGCATATCA	GTGGCAAATT	GA CT TAAAAT	CCATACCCCT	ACTATTTTAA	GACCATTGTC	120
CTTTGGAGCA	GAGAGACAGA	CTCTCCCATI	GAGAGGICTT	GCTATAAGCC	TTCATCCGGA	180
GAGTGTAGGG	TAGAGGGCCT	GGGT TAAAGT	TGCA GATTAC	TGCAGTGATT	TTACATGTAA	240
ATGTCCATTT	TAGATCAACT	GGAATGGATG	GTACAGCTGT	GTGGTGCTTC	TGTGGTGAAG	300
GAGCTTTTCAT	CATTACCCCT	TGGCACAGTA	AGTATTGGGT	GCCCIGTCAG	TGTGGGAGGA	360
CACAATATTIC	TCTCCTGTGA	GCAAAGACTGG	CACCTGT CAG	TCCCTATGGA	TGCCCTACT	420
GTAGCCTCAG	AAGTCTTCTC	TGCCCACATA	CTGTGCCAA	AAGACTCCAT		470

(2) INFORMATION FOR SEO ID NO:33:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 517 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(i v) ANTI-SENSE: NO

(v i) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:33:

GGTGOTACGT	GTCGTGTAOTT	CCAGCTACTT	GGGAGGCTGA	GATGGAAGGA	TTGCTTGAGC	60
CCAGGAGGCA	GAGGTGGNAN	NTTACGCTGA	GATCACACCA	CTGCACTCCA	GCCTGGGTGA	120
CAGAGCAAGA	CCCTGTCTCA	AAAACAAACA	AAAAAAATGA	TGAAGTGACA	GTTCAGTAG	180
TCCTACTTTG	ACACTTTGAA	TGCTCTTTCC	TTCTGGGGG	TCCAGGGTGT	CCACCCAATT	240
GTGGTTGTGC	AGCCAGATGC	CTGGACAGAG	GACAATGGCT	TCCATGGTAA	GGTGCCTCGC	300
ATGTACCTGT	GCTATTAGTG	GGGTCCTTGT	GCATGGGTTT	GGTTTATCAC	TCATTACCTG	360
GTGCTTGAGT	AGCACAGTTC	TTGGCACATT	TTTAAATATT	TGTTGAATGA	ATGGCTAAAA	420
TGTCITTTTG	ATGTTTTTAT	TGTTATTTGT	TTTATATTGT	AAAAGTAATA	CATGAACTGT	480
TTCCATGGGG	TGGGAGTAAG	ATATGAATGT	TCATCAC			517

(2) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 434 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(i v) ANTI-SENSE: NO

(v i) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:34:

CAGTAATCCT NAGAACTCAT ACGACCGGGC CCCTGGAGTC GNTGNTTNGA GCCTAGTCCN 60

5,753,441

129

130

-continued

GGAGAATGAA TTGACACTAA TCTCTGCTTG TGTTCTCTGT CTCCAGCAAT TGGGCAGATG 120
TGTGAGGCAC CTGTGGTGAC CCGAGAGTGG GTGTTGGACA GTGTAGCACT CTACCAGTGC 180
CAGGAGCTGG ACACCTACCT GATACCCAG ATCCCCACA GCCACTACTG ACTGCAGCCA 240
GCCACAGGTA CAGAGCCACA GGACCCCAAG AATGAGCTTA CAAAGTGGCC TTTCCAGGCC 300
CTGGGAGCTC CTCTCACTCT TCAATCCTTC TACTGTCCTG GCTACTAAAT ATTTTATGTA 360
CATCAGCCTG AAAAGGACTT CTGGCTATGC AAGGGTCCCT TAAAGATTTT CTGCTTGAAG 420
TCTCCCTTGG AAAT 434

(2) INFORMATION FOR SEQ ID NO:35:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 30 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:

- (A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:35:

GATAAATTAA AACTGCGACT GCGCGGCGTG 30

(2) INFORMATION FOR SEQ ID NO:36:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 30 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:

- (A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:36:

GTAGTAGAGT CCCGGGAAAG GGACAGGGGG 30

(2) INFORMATION FOR SEQ ID NO:37:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 30 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:

- (A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:37:

ATATATATAT GTTTTCTAA TGTGTTAAAG 30

(2) INFORMATION FOR SEQ ID NO:38:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 30 base pairs

5,753,441

131

132

-continued

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: *Homo sapiens*

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:38:

GTAAGTCAGC ACAAGAGTGT ATTAATTITGG

30

(2) INFORMATION FOR SEQ ID NO:39:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:39:

TTTCTTTTTC TCCCCCCCCT ACCCTGCTAG

30

(2) INFORMATION FOR SEQ ID NO:40:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:40:

GTAAGTTTGA ATGTGTTATG TGGCTCCATT

30

(2) INFORMATION FOR SEQ ID NO:41:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: *Homo sapiens*

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:41:

AGCTACTTTT TTTTTTTTTT TTTGAGACAG

30

(2) INFORMATION FOR SEQ ID NO:42:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs

5.753.441

133

134

-continued

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:42:

GTAAGTGCAC ACCACCATAT CCAGCTAAAT

(2) INFORMATION FOR SEQ ID NO:43:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:43:

AATTGTTCTT TCTTCTTTA TAATTTATAG

(2) INFORMATION FOR SEQ ID NO:44:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:44:

GTATATAATT TGGTAATGAT GCTAGGTTGG

(2) INFORMATION FOR SEQ ID NO:45:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:45:

GAGTGTGTTT CTCAAACAAT TTAATTTCAAG

(2) INFORMATION FOR SEQ ID NO:46:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs

A000337

5,753,441

135

136

-continued

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:46:

GTAAAGTGTTC AATATCCCAA GAATGACACT 3 0

(2) INFORMATION FOR SEQ ID NO:47:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:47:

AAACATAATG TTTTCCTTG TATTTTACAG 3 0

(2) INFORMATION FOR SEQ ID NO:48:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:48:

GTAAACCAT TGTTCCTT CTCTTCTTC 3 0

(2) INFORMATION FOR SEQ ID NO:49:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:49:

TGCTTGACTG TCTTTACCA TACTGTTTAG 3 0

(2) INFORMATION FOR SEQ ID NO:50:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs

5,753,441

137

138

-continued

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:50:

GTAAAGGGTCT CAGGTTTTTT AAGTATTTAA 3 0

(2) INFORMATION FOR SEQ ID NO:51:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:51:

TGATTTATTT TTTGGGGGGA AATTTTTTAG 3 0

(2) INFORMATION FOR SEQ ID NO:52:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:52:

GTGAGTCAAA GAGAACCCTT GTCTATGAAG 3 0

(2) INFORMATION FOR SEQ ID NO:53:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:53:

TCTTATTAGG ACTCTGCTT TTCCCTATAG 3 0

(2) INFORMATION FOR SEQ ID NO:54:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs

5,753,441

139

140

-continued

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:54:

GTAAATGGCAA AGTTTGCCAA CTTAACAGGC 3 0

(2) INFORMATION FOR SEQ ID NO:55:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:55:

GAGTACCTTG TTATTTTGT ATATTTTCAG 3 0

(2) INFORMATION FOR SEQ ID NO:56:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:56:

GTATTGGAAC CAGGTTTTTG TGTTCGCC 3 0

(2) INFORMATION FOR SEQ ID NO:57:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:57:

ACATCTGAAC CTCTGTTTTT GTATTTTAAG 3 0

(2) INFORMATION FOR SEQ ID NO:58:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs

5,753,441

141

142

-continued

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:58:

AGGTAAAAAG CGTGTGTGTG TGTGCACATG 3 0

(2) INFORMATION FOR SEQ ID NO:59:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:59:

CATTTTCTTG GTACCATTTA TCGTTTTTGA 3 0

(2) INFORMATION FOR SEQ ID NO:60:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:60:

GTGTGTATTG TTGGCCAAAC ACTGATATCT 3 0

(2) INFORMATION FOR SEQ ID NO:61:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:61:

AGTAGATTTG TTTTCTCATT CCATTTAAAG 3 0

(2) INFORMATION FOR SEQ ID NO:62:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs

5,753,441

143

144

-continued

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:62:

GTAAGAAACA TCAATGTAAA GATGCTGTGG 3 0

(2) INFORMATION FOR SEQ ID NO:63:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:63:

ATGGTTTTCT CCTTCATTT ATCTTTCTAG 3 0

(2) INFORMATION FOR SEQ ID NO:64:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:64:

GTAATATTTT ATCTGCTGTA TTGGAACAAA 3 0

(2) INFORMATION FOR SEQ ID NO:65:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:65:

TGTAAATTAA ACTTCICCCA TTCCTTTTCA 3 0

(2) INFORMATION FOR SEQ ID NO:66:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs

5,753,441

145

146

-continued

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:66:

GTGAGTGTAT CCATATGTAT CTCCTAATG 3 0

(2) INFORMATION FOR SEQ ID NO:67:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:67:

ATGATAATGG AATATTGAT TTAATTTCA 3 0

(2) INFORMATION FOR SEQ ID NO:68:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:68:

GTATACCAAG AACCTTTACA GAATACCTTG 3 0

(2) INFORMATION FOR SEQ ID NO:69:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:69:

CTAATCCTTT GAGTGTTTTT CATTCTGCA 3 0

(2) INFORMATION FOR SEQ ID NO:70:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs

5,753,441

147

148

-continued

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:70:

GTAAGTATAA TACTATTTCT CCCCTCCTCC 3 0

(2) INFORMATION FOR SEQ ID NO:71:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:71:

TGTAACCTGT CTTTCTATG ATCTCTTAG 3 0

(2) INFORMATION FOR SEQ ID NO:72:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:72:

GTAAGTACTT GATGTTACAA ACTAACCAGA 3 0

(2) INFORMATION FOR SEQ ID NO:73:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:73:

TCCTGATGGG TTGTGTTTGG TTTCTTTCAG 3 0

(2) INFORMATION FOR SEQ ID NO:74:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs

5,753,441

149

150

-continued

```
( B ) TYPE: nucleic acid
( C ) STRANDEDNESS: single
( D ) TOPOLOGY: linear

( i i ) MOLECULE TYPE: DNA (genomic)

( i i i ) HYPOTHETICAL: NO

( v i ) ORIGINAL SOURCE:
  ( A ) ORGANISM: Homo sapiens

( x i ) SEQUENCE DESCRIPTION: SEQ ID NO:74:

GTAAAGCTCC CTCCTCAAG TTGACAAAAA 30

( 2 ) INFORMATION FOR SEQ ID NO:75:

  ( i ) SEQUENCE CHARACTERISTICS:
    ( A ) LENGTH: 30 base pairs
    ( B ) TYPE: nucleic acid
    ( C ) STRANDEDNESS: single
    ( D ) TOPOLOGY: linear

  ( i i ) MOLECULE TYPE: DNA (genomic)

  ( i i i ) HYPOTHETICAL: NO

  ( v i ) ORIGINAL SOURCE:
    ( A ) ORGANISM: Homo sapiens

  ( x i ) SEQUENCE DESCRIPTION: SEQ ID NO:75:

CTGTCCCTCT CTCTTCCTCT CTTCTTCCAG 30

( 2 ) INFORMATION FOR SEQ ID NO:76:

  ( i ) SEQUENCE CHARACTERISTICS:
    ( A ) LENGTH: 30 base pairs
    ( B ) TYPE: nucleic acid
    ( C ) STRANDEDNESS: single
    ( D ) TOPOLOGY: linear

  ( i i ) MOLECULE TYPE: DNA (genomic)

  ( i i i ) HYPOTHETICAL: NO

  ( v i ) ORIGINAL SOURCE:
    ( A ) ORGANISM: Homo sapiens

  ( x i ) SEQUENCE DESCRIPTION: SEQ ID NO:76:

GTAAGAGCCT GGGAGAACCC CAGAGTTCCA 30

( 2 ) INFORMATION FOR SEQ ID NO:77:

  ( i ) SEQUENCE CHARACTERISTICS:
    ( A ) LENGTH: 30 base pairs
    ( B ) TYPE: nucleic acid
    ( C ) STRANDEDNESS: single
    ( D ) TOPOLOGY: linear

  ( i i ) MOLECULE TYPE: DNA (genomic)

  ( i i i ) HYPOTHETICAL: NO

  ( v i ) ORIGINAL SOURCE:
    ( A ) ORGANISM: Homo sapiens

  ( x i ) SEQUENCE DESCRIPTION: SEQ ID NO:77:

AGTGATTTTA CATGTAAATG TCCATTTTAG 30

( 2 ) INFORMATION FOR SEQ ID NO:78:

  ( i ) SEQUENCE CHARACTERISTICS:
    ( A ) LENGTH: 30 base pairs
```

5,753,441

151

152

-continued

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: *Homo sapiens*

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:78:

GTAAGTATTG GGTGCCCTGT CAGTGTGGGA

30

(2) INFORMATION FOR SEO ID NO:79:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:79:

TTGAATGCTC TTTCCTTCCT GGGGATCCAG

30

(2) INFORMATION FOR SEQ ID NO:80:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: *Homo sapiens*

(x i) SEQUENCE DESCRIPTION: SBO ID NO:80:

GTAAAGGTGCC TCGCATGTAC CTGTGCTATT

30

(2) INFORMATION FOR SEQ ID NO:81:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(i i i) HYPOTHETICAL: NO

(v i) ORIGINAL SOURCE:
(A) ORGANISM: *Homo sapiens*

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:81:

CTAATCTCTG CTTGTGTTCT CTGTCTCCAG

30

(2) INFORMATION FOR SEQ ID NO:82:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 42 amino acids

5,753,441

153

154

-continued

(B) TYPE: amino acid	
(C) STRANDEDNESS:	
(D) TOPOLOGY: linear	
(i i) MOLECULE TYPE: peptide	
(i i i) HYPOTHETICAL: NO	
(v i) ORIGINAL SOURCE:	
(A) ORGANISM: Homo sapiens	
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:82:	
Cys	Pro Ile Cys Leu Glu Leu Ile Lys Glu Pro Val Ser Thr Lys Cys
1	5 10 15
Asp His Ile Phe Cys Lys Phe Cys Met Leu Lys Leu Leu Asn Gln Lys	
	20 25 30
Lys Gly Pro Ser Gln Cys Pro Leu Cys Lys	
	35 40
(2) INFORMATION FOR SEQ ID NO:83:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 45 amino acids	
(B) TYPE: amino acid	
(C) STRANDEDNESS:	
(D) TOPOLOGY: linear	
(i i) MOLECULE TYPE: peptide	
(i i i) HYPOTHETICAL: NO	
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:83:	
Cys	Pro Ile Cys Leu Glu Leu Leu Lys Glu Pro Val Ser Ala Asp Cys
1	5 10 15
Asn His Ser Phe Cys Arg Ala Cys Ile Thr Leu Asn Tyr Glu Ser Asn	
	20 25 30
Arg Asn Thr Asp Gly Lys Gly Asn Cys Pro Val Cys Arg	
	35 40 45
(2) INFORMATION FOR SEQ ID NO:84:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 41 amino acids	
(B) TYPE: amino acid	
(C) STRANDEDNESS:	
(D) TOPOLOGY: linear	
(i i) MOLECULE TYPE: peptide	
(i i i) HYPOTHETICAL: NO	
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:84:	
Cys	Pro Ile Cys Leu Asp Met Leu Lys Asn Thr Met Thr Thr Lys Glu
1	5 10 15
Cys Leu His Arg Phe Cys Ser Asp Cys Ile Val Thr Ala Leu Arg Ser	
	20 25 30
Gly Asn Lys Glu Cys Pro Thr Cys Arg	
	35 40
(2) INFORMATION FOR SEQ ID NO:85:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 42 amino acids	
(B) TYPE: amino acid	
(C) STRANDEDNESS:	
(D) TOPOLOGY: linear	
(i i) MOLECULE TYPE: peptide	

5,753,441

155

156

-continued

(i i i) HYPOTHETICAL: NO

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:85:

Cys	Pro	Val	Cys	Leu	Gln	Tyr	Phe	Ala	Glu	Pro	Met	Met	Leu	Asp	Cys
1				5					10					15	
Gly	His	Asn	Ile	Cys	Cys	Ala	Cys	Leu	Ala	Arg	Cys	Trp	Gly	Thr	Ala
			20					25					30		
Cys	Thr	Asn	Val	Ser	Cys	Pro	Gln	Cys	Arg						
			35				40								

What is claimed is:

1. A method for screening germline of a human subject for an alteration of a BRCA1 gene which comprises comparing germline sequence of a BRCA1 gene or BRCA1 RNA from a tissue sample from said subject or a sequence of BRCA1 cDNA made from mRNA from said sample with germline sequences of wild-type BRCA1 gene, wild-type BRCA1 RNA or wild-type BRCA1 cDNA, wherein a difference in the sequence of the BRCA1 gene, BRCA1 RNA or BRCA1 cDNA of the subject from wild-type indicates an alteration in the BRCA1 gene in said subject.

2. The method of claim 1 wherein the wild-type BRCA1 gene has the sequence set forth in SEQ ID NO:1.

3. The method of claim 1 wherein the nucleic acid sequence of BRCA1 RNA from the subject is compared to nucleic acid sequences of wild-type BRCA1 gene, BRCA1 RNA or BRCA1 cDNA.

4. The method of claim 3 wherein the nucleic acid sequence is compared by hybridizing a BRCA1 gene probe which specifically hybridizes to a BRCA1 allele to RNA isolated from said subject and detecting of the presence of a hybridization product, wherein a presence of said product indicates the presence of said allele in the subject.

5. The method of claim 1 wherein a regulatory region of the BRCA1 gene from said subject is compared with a regulatory region of wild-type BRCA1 gene sequences, said regulatory region corresponding to nucleotides 1-1531 of SEQ ID NO:13.

6. The method of claim 1 wherein a germline nucleic acid sequence is compared by obtaining a first BRCA1 gene fragment from a BRCA1 gene from a human sample and a second BRCA1 gene fragment from a wild-type BRCA1 gene, said second fragment corresponding to said first fragment forming single-stranded DNA from said first BRCA1 gene fragment and from said second BRCA1 gene fragment, electrophoresing said single-stranded DNAs on a non-denaturing polyacrylamide gel, comparing the mobility of said single-stranded DNAs on said gel to determine if said single-stranded DNA from said first BRCA1 gene fragment is shifted relative to said second BRCA1 gene fragment and sequencing said single-stranded DNA from said first BRCA1 gene fragment having a shift in electrophoretic mobility.

7. The method of claim 1 wherein a germline nucleic acid sequence is compared by hybridizing a BRCA1 gene probe which specifically hybridizes to a BRCA1 allele to genomic DNA isolated from said sample and detecting the presence of a hybridization product wherein a presence of said product indicates the presence of said allele in the subject.

8. The method of claim 1 wherein a germline nucleic acid sequence is compared by amplifying all or part of a BRCA1 gene from said sample using a set of primers to produce amplified nucleic acids and sequencing the amplified nucleic acids.

9. The method of claim 1 wherein a germline nucleic acid sequence is compared by amplifying all or part of a BRCA1 gene using a primer specific for a specific BRCA1 mutant allele and detecting the presence of an amplified product, wherein the presence of said product indicates the presence of said specific allele.

10. The method of claim 1 wherein a germline nucleic acid sequence is compared by molecularly cloning all or part of a BRCA1 gene from said sample to produce a cloned nucleic acid and sequencing the cloned nucleic acid.

11. The method of claim 1 wherein a germline nucleic acid sequence is compared by obtaining a first BRCA1 gene fragment from (a) BRCA1 gene genomic DNA isolated from said sample, (b) BRCA1 RNA isolated from said sample or (c) BRCA1 CDNA made from mRNA isolated from said sample, obtaining a second BRCA1 gene fragment from (a) wild-type BRCA1 genomic DNA, (b) wild-type BRCA1 RNA or (c) wild-type cDNA made from wild-type mRNA, said second BRCA1 gene fragment corresponding to said first BRCA1 gene fragment, forming single-stranded DNA from said first BRCA1 gene fragment and from said second BRCA1 gene fragment, forming a heteroduplex consisting of single-stranded DNA from said BRCA1 gene fragment and single-stranded DNA from said second BRCA1 gene fragment, analyzing the heteroduplex to determine if said single-stranded DNA from said first BRCA1 gene fragment has a mismatch relative to said single-stranded DNA from said second BRCA1 gene fragment and sequencing said first single-stranded DNA from said first BRCA1 gene fragment having a mismatch.

12. The method of claim 1 wherein a germline nucleic acid sequence is compared by amplifying BRCA1 nucleic acids from said sample to produce amplified nucleic acids, hybridizing the amplified nucleic acids to a BRCA1 DNA probe specific for a BRCA1 allele and detecting the presence of a hybridization product, wherein the presence of said product indicates the presence of said allele in the subject.

13. The method of claim 1 wherein a germline nucleic acid sequence is compared by analyzing BRCA1 nucleic acids in said sample for a deletion mutation.

14. The method of claim 1 wherein a germline nucleic acid sequence is compared by analyzing BRCA1 nucleic acids in said sample for a point mutation.

15. The method of claim 1 wherein a germline nucleic acid sequence is compared by analyzing BRCA1 nucleic acids in said sample for an insertion mutation.

16. The method of claim 1 wherein a germline nucleic acid sequence is compared by hybridizing the tissue sample in situ with a nucleic acid probe specific for a BRCA1 allele and detecting the presence of a hybridization product, wherein a presence of said product indicates the presence of said allele in the subject.

17. The method of claim 1 wherein a nucleic acid of a germline BRCA1 cDNA made from mRNA from said

A000348

157

sample is compared to nucleic acid sequences of wild-type BRCA1 gene, BRCA1 RNA or BRCA1 cDNA.

18. The method of claim 1 wherein a nucleic acid sequence of a germline BRCA1 gene from the subject is compared to nucleic acid sequences of wild-type BRCA1 gene, BRCA1 RNA or BRCA1 cDNA.

19. The method of claim 1 wherein said difference is selected from the group consisting of missense mutations within the zinc finger motif, deletions, insertions, frameshift mutations, nonsense mutations and splice site mutations.

20. A method for detecting a germline alteration in a BRCA1 gene, said alteration selected from the group consisting of the alterations set forth in Tables 11 and 12 which comprises analyzing a sequence of the BRCA1 gene or BRCA1 RNA from a human sample or analyzing the sequence of BRCA1 CDNA made from mRNA from said sample.

21. The method of claim 20 wherein a germline alteration is detected by hybridizing a BRCA1 gene probe which specifically hybridizes to an allele of one of said alterations to RNA isolated from said human sample and detecting the presence of a hybridization product, wherein the presence of said product indicates the presence of said allele in the sample.

22. The method of claim 20 wherein a germline alteration is detected by hybridizing a BRCA1 gene probe which specifically hybridizes to one of said alterations to genomic DNA isolated from said sample and detecting the presence of a hybridization product, wherein the presence of said product indicates the presence of said alteration in the sample.

23. The method of claim 20 wherein a germline alteration is detected by amplifying all or part of a BRCA1 gene in said sample using a set of primers to produce amplified nucleic acids and sequencing the amplified nucleic acids.

24. The method of claim 20 wherein a germline alteration is detected by amplifying part of a BRCA1 gene in said sample using a primer specific for an allele having one of said alterations and detecting the presence of an amplified product, wherein the presence of said product indicates the presence of said allele in the sample.

25. The method of claim 20 wherein a germline alteration is detected by molecularly cloning all or part of a BRCA1 gene in said sample to produce a cloned nucleic acid and sequencing the cloned nucleic acid.

26. The method of claim 20 wherein a germline alteration is detected by amplifying BRCA1 gene nucleic acids in said sample, hybridizing the amplified nucleic acids to a BRCA1 DNA probe specific for one of said alterations and detecting

158

the presence of a hybridization product, wherein the presence of said product indicates the presence of said alteration.

27. A method for screening for a germline alteration in a BRCA1 gene in a human subject which comprises analyzing a BRCA1 polypeptide from a tissue sample from said subject for an altered BRCA1 polypeptide by (i) detecting either a full length BRCA1 polypeptide or a truncated BRCA1 polypeptide or (ii) contacting an antibody which specifically binds to an epitope of an altered BRCA1 polypeptide to the BRCA1 polypeptide from said sample and detecting bound antibody, wherein the presence of a truncated protein or bound antibody indicates the presence of a germline alteration in the BRCA1 gene.

28. The method of claim 27 wherein a BRCA1 polypeptide is analyzed by detecting a truncated BRCA1 polypeptide.

29. The method of claim 27 wherein a BRCA1 polypeptide is analyzed by contacting an antibody which specifically binds to an epitope of an altered BRCA1 polypeptide to the BRCA1 polypeptide from said sample.

30. The method of claim 20 wherein said alteration consists of a deletion of 11 nucleotides corresponding to base numbers 189-199 in SEQ ID NO:1.

31. The method of claim 21 wherein said alteration consists of a deletion of 11 nucleotides corresponding to base numbers 189-199 in SEQ ID NO:1.

32. The method of claim 22 wherein said alteration consists of a deletion of 11 nucleotides corresponding to base numbers 189-199 in SEQ ID NO:1.

33. The method of claim 23 wherein said alteration consists of a deletion of 11 nucleotides corresponding to base numbers 189-199 in SEQ ID NO:1.

34. The method of claim 24 wherein said alteration consists of a deletion of 11 nucleotides corresponding to base numbers 189-199 in SEQ ID NO:1.

35. The method of claim 25 wherein said alteration consists of a deletion of 11 nucleotides corresponding to base numbers 189-199 in SEQ ID NO:1.

36. The method of claim 26 wherein said alteration consists of a deletion of 11 nucleotides corresponding to base numbers 189-199 in SEQ ID NO:1.

37. A kit for screening for an alteration in a BRCA1 gene in a human subject which comprises at least one antibody (i) which specifically binds to wild-type BRCA1 polypeptide but not a truncated BRCA1 polypeptide or (ii) which specifically binds to an epitope of an altered BRCA1 polypeptide.

* * * * *

UNITED STATES PATENT AND TRADEMARK OFFICE
CERTIFICATE OF CORRECTION

PATENT NO. : 5,753,441
DATED : May 19, 1998
INVENTOR(S) : Mark H. Skolnick et al.

Page 1 of 1

It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

Title page,

Item [22], "Jan. 5, 1996" should be -- June 7, 1995 --.

Signed and Sealed this

Thirtieth Day of July, 2002

Attest:



Attesting Officer

JAMES E. ROGAN
Director of the United States Patent and Trademark Office

A000350

UNITED STATES PATENT AND TRADEMARK OFFICE
CERTIFICATE OF CORRECTION

PATENT NO. : 5,753,441
DATED : May 19, 1998
INVENTOR(S) : Mark H. Skolnick et al.

Page 1 of 1

It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

Title page,

Item [22], "Jan. 5, 1996" should be -- June 7, 1995 --.

Signed and Sealed this

Thirtieth Day of July, 2002

Attest:



Attesting Officer

JAMES E. ROGAN
Director of the United States Patent and Trademark Office

A000351



US006033857A

United States Patent [19]
Tavtigian et al.

[11] **Patent Number:** **6,033,857**
[45] **Date of Patent:** **Mar. 7, 2000**

- [54] **CHROMOSOME 13-LINKED BREAST CANCER SUSCEPTIBILITY GENE**
- [75] Inventors: **Sean V. Tavtigian; Alexander Kamb**, both of Salt Lake City, Utah; **Jacques Simard**, St. Augustin de Desmuures, Canada; **Fergus Couch**, St. Davids, Pa.; **Johanna M. Rommens**, Toronto, Canada; **Barbara L. Weber**, Merion, Pa.
- [73] Assignees: **Myriad Genetics, Inc.**, Salt Lake City, Utah; **Endo Recherche, Inc.; HSC Research & Development Limited Parntership**, both of Canada; **Trustees of the Univ. of Pennsylvania**, Philadelphia, Pa.
- [21] Appl. No.: **09/044,946**
- [22] Filed: **Mar. 20, 1998**

Related U.S. Application Data

- [60] Division of application No. 08/639,501, Apr. 29, 1996, Pat. No. 5,837,492, which is a continuation-in-part of application No. 08/585,391, Jan. 11, 1996, abandoned, which is a continuation-in-part of application No. 08/576,559, Dec. 21, 1995, abandoned, which is a continuation-in-part of application No. 08/575,359, Dec. 20, 1995, abandoned, which is a continuation-in-part of application No. 08/573,779, Dec. 18, 1995, abandoned.
- [51] **Int. Cl.⁷** **C07H 21/00**; C12N 15/63; C12N 15/79; C12N 15/11; C12N 15/09
- [52] **U.S. Cl.** **435/6**; 435/7.2; 435/69.1; 435/325; 435/320.1; 536/23.1; 536/23.5
- [58] **Field of Search** 435/6, 7.2, 69.1, 435/325, 320.1; 536/23.1, 23.5

References Cited

FOREIGN PATENT DOCUMENTS

2307477	5/1997	United Kingdom .
9515334	6/1995	WIPO .
9519369	7/1995	WIPO .
9719110	5/1997	WIPO .

OTHER PUBLICATIONS

"Isolation of expressed sequences that include a gene for familial breast cancer (BRCA2) and other novel transcripts from a five megabase region on chromosome 13q12," ANK Jacob et al., *Oncogene* (1996) 13, pp. 213–221.

"Generation of an Integrated Transcription Map of the BRCA2 Region on Chromosome 13q12–q13," Fergus J. Couch et al., *Genomics* 36, Article No. 0428 (1996), pp. 86–99.

"A Strong Candidate for the Breast and Ovarian Cancer Susceptibility Gene BRCA1," Yoshiio Miki et al., *Science*, vol. 266, Oct. 7, 1994, pp. 66–71.

"Different Tumor types from BRCA2 Carriers Show Wild-Type Chromosome Deletions on 13q12–q13," Julius Gudmundsson et al., *Cancer Research* 55, Nov. 1, 1995, pp. 4830–4832.

"Somatic and Germline Mutations of the BRCA2 Gene in Sporadic Ovarian Cancer¹," Karen A. Foster et al., *Cancer Research* 56, Aug. 15, 1996, pp. 3622–3625.

"A Common Mutation in BRCA2 That Predisposes to a Variety of Cancers Is Found in Both Jewish Ashkenazi and Non-Jewish Individuals¹," David B. Berman et al., *Cancer Research* 56, Aug. 1, 1996, pp. 3409–3414.

"Patterns of Loss of Heterozygosity at Loci From Chromosome Arm 13q Suggest a Possible Involvement of BRCA2 in Sporadic Breast Tumors," Fabienne Kerangueven et al., *Gene, Chromosomes on Cancer* 13, (1995), p. 291–294.

"Loss of heterozygosity in sporadic breast tumours at the BRCA2 locus on chromosome 13q12–q13," A-M Cleton-Jansen, et al., *British Journal of Cancer* (1995) 72, pp. 1241–1244.

"Consistent loss of the wild type allele in breast cancers from a family linked to the BRCA2 gene on chromosome 13q12–13," Nadine Collins et al., *Short Report*, Revised Jan. 30, 1995, accepted Jan. 30, 1995, pp. 1673–1675.

"Loss of Heterozygosity on Chromosome 13 is Common Only in the Biologically More Aggressive Subtypes of Ovarian Epithelial Tumors and Is Associated with Normal Retinoblastoma Gene Expression¹," Timothy M. Kim et al., *Advances in Brief*, Accepted Dec. 17, 1993, pp. 605–609.

"BRCA2 germline mutations in male breast cancer cases and breast cancer families," Fergus J. Couch et al., *Nature Genetics*, vol. 13, May 1996, pp. 123–125.

"Mutation analysis of the BRCA2 gene in 49 site-specific breast cancer families," Catherine M. Phelan et al., *Nature Genetics*, vol. 13, May 1996, pp. 120–122.

"Recurrent BRCA2 6174delT mutations in Ashkenazi Jewish woman affected by breast cancer," Susan Neuhausen et al., *Nature Genetics*, vol. 13, May 1996, pp. 126–128.

"Loss of heterozygosity in human ductal breast tumors indicated a recessive mutation on chromosome 13," Catharina Lundberg et al., *Proc. Natl. Acad. Sci. USA*, vol. 84, Apr. 1987, pp. 2372–2376.

(List continued on next page.)

Primary Examiner—Karen M. Hauda
Attorney, Agent, or Firm—Rothwell, Figg, Ernst & Kurz, P.C.

[57] **ABSTRACT**

The present invention relates generally to the field of human genetics. Specifically, the present invention relates to methods and materials used to isolate and detect a human breast cancer predisposing gene (BRCA2), some mutant alleles of which cause susceptibility to cancer, in particular breast cancer. More specifically, the invention relates to germline mutations in the BRCA2 gene and their use in the diagnosis of predisposition to breast cancer. The present invention further relates to somatic mutations in the BRCA2 gene in human breast cancer and their use in the diagnosis and prognosis of human breast cancer. Additionally, the invention relates to somatic mutations in the BRCA2 gene in other human cancers and their use in the diagnosis and prognosis of human cancers. The invention also relates to the therapy of human cancers which have a mutation in the BRCA2 gene, including gene therapy, protein replacement therapy and protein mimetics. The invention further relates to the screening of drugs for cancer therapy. Finally, the invention relates to the screening of the BRCA2 gene for mutations, which are useful for diagnosing the predisposition to breast cancer.

8 Claims, 9 Drawing Sheets

6,033,857

Page 2

OTHER PUBLICATIONS

"Identification by representational difference analysis of a homozygous deletion in pancreatic carcinoma that lies within the BRCA2 region," Mieke Schuttelle et al., *Proc. Natl. Acad. Sci. USA*, vol. 92, Jun. 1995, pp. 5950-5954.

"Linkage to BRCA2 region in hereditary male breast cancer," Steinunn Thorlacius et al., *The Lancet*, vol. 346, No. 8974, Aug. 26, 1995, pp. 544-545.

"Localization of a Breast Cancer Susceptibility Gene, BRCA2, to Chromosome 13q12-13," Richard Wooster et al., *Science*, vol. 265, Sep. 30, 1994, pp. 2088-2090.

"Identification of the breast cancer susceptibility gene BRCA2," Richard Wooster et al., *Nature*, vol. 378, Dec. 21/28, 1995, pp. 789-792.

"Regulating gene expression in transgenic animals," Catherine A. Kappel et al., *Biotechnology* 1992, 3, pp. 548-553.

"The complete BRCA2 gene and mutations in chromosome 13q-linked kindreds," S. V. Tavtigian et al., *Nature Genetics*, vol. 12, Mar. 1996, pp. 333-337.

"Confirmation of a susceptibility locus on chromosome 13 in Australian breast cancer families," Sean M. Grimmond et al., *Hum Genet* (1996) 98, pp. 80-85.

"Further enigmatic variations," Kevin Davies, *Nature*, vol. 378, Dec. 21/28, 1995, pp. 762–763.

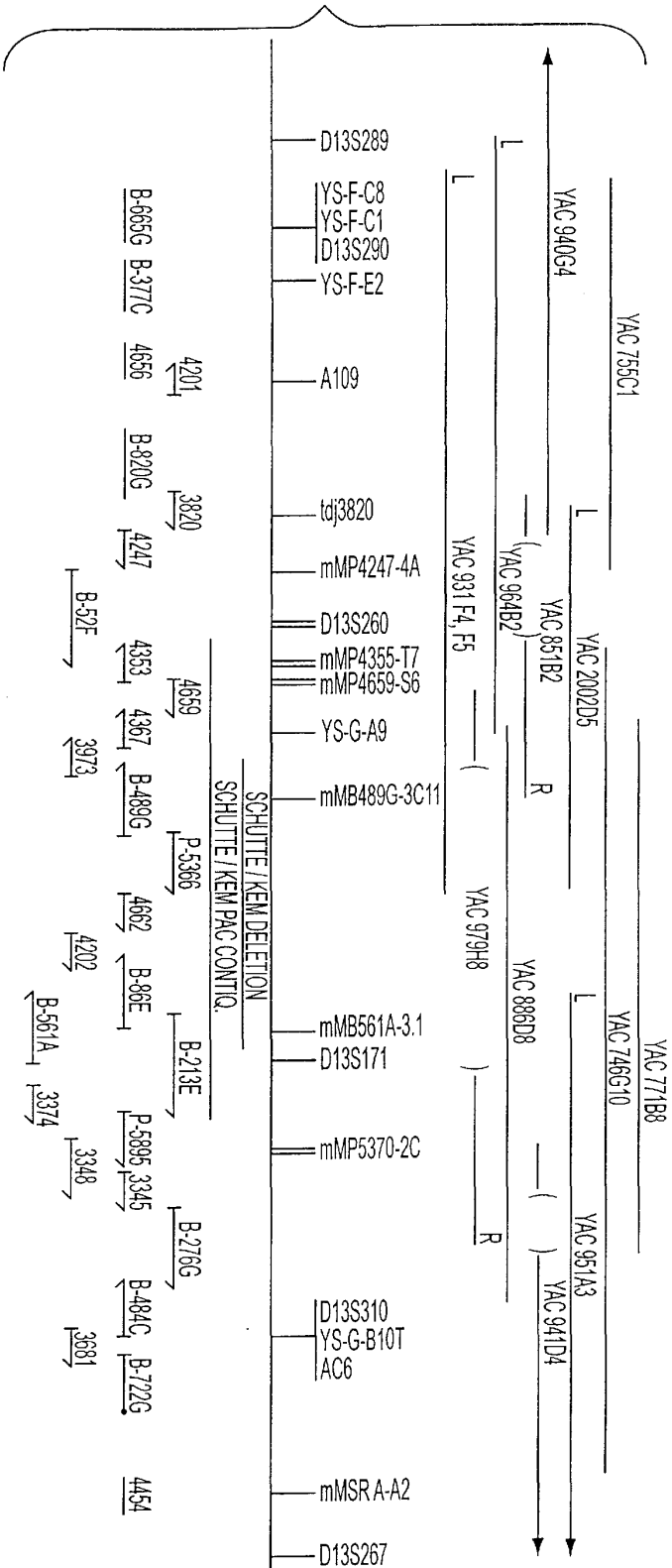
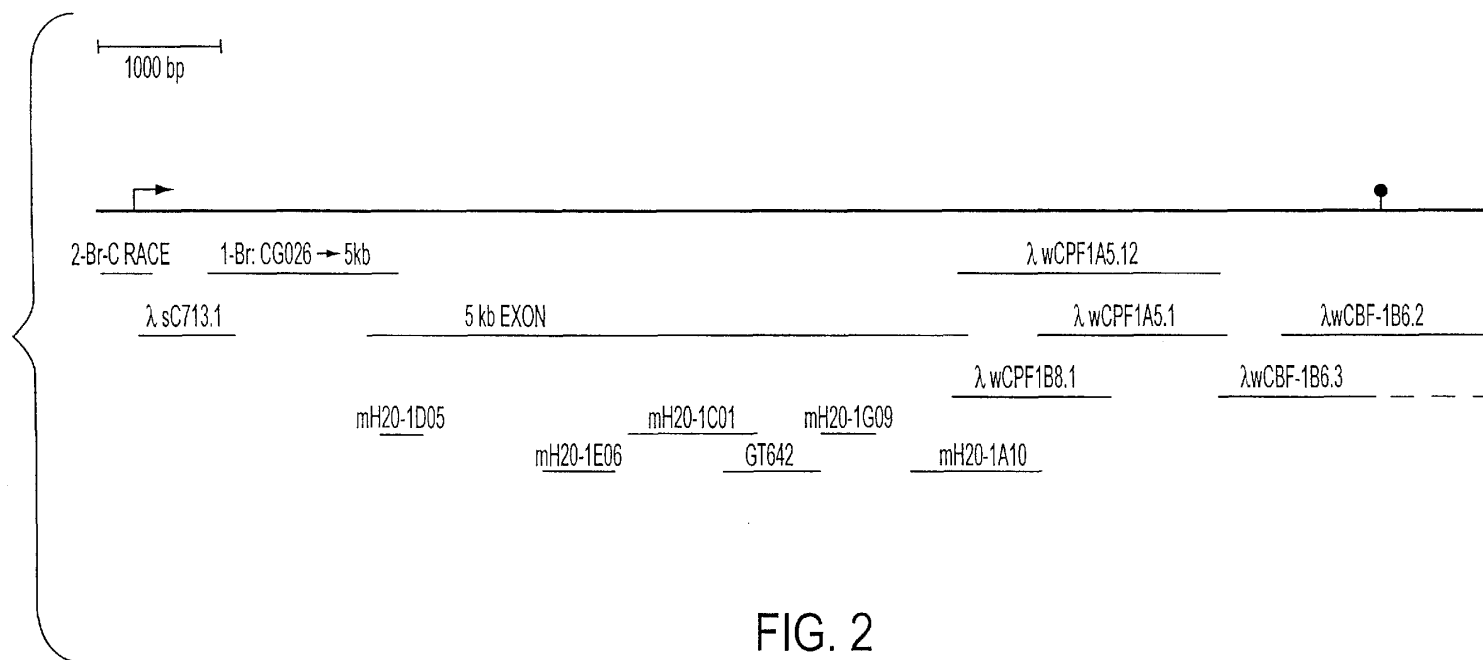


FIG. 1



U.S. Patent

Mar. 7, 2000

Sheet 2 of 9

6,033,857

A000450

6,033,857

FIG. 3A

A000451

6,033,857

FIG. 3B

A000452

6541	ATACTTCCTC	GTGTTGATAA	GAGAAACCCA	GAGCACTGTG	TAAACTCAGA	AATGGAAAAA
6601	ACCTGCAGTA	AAGAATTTAA	ATTATCAAAT	AACTTAAATG	TTGAAGGTGG	TTCTTCAGAA
6661	AATAATCACT	CTATTAAAGT	TTCTCCATAT	CTCTCTCAAT	TTCAACAAGA	CAAACAACAG
6721	TTGGTATTAG	GAACCAAAGT	CTCACTTGTT	GAGAACATTC	ATGTTTTGGG	AAAAGAACAG
6781	GCTTCACCTA	AAAACGTAAA	AATGGAAATT	GGTAAAACTG	AAACTTTTTT	TGATGTTCC
6841	GTGAAAACAA	ATATAGAAGT	TTGTTCTACT	TACTCCAAAG	ATTCAAGAAA	CTACTTTGAA
6901	ACAGAAGCAG	TAGAAATTGC	TAAAGCTTTT	ATGGAAGATG	ATGAACTGAC	AGATTCTAAA
6961	CTGCCAAGTC	ATGCCACACA	TTCTCTTTTT	ACATGTCCCG	AAAATGAGGA	AATGGTTTTG
7021	TCAAATTCAA	GAATTGGAAA	AAGAAGAGCA	GAGCCCTTAA	TCTTAGTGGG	AGAACCCTCA
7081	ATCAAAAGAA	ACTTATTAAA	TGAATTTGAC	AGGATAATAG	AAAATCAAGA	AAAATCCTTA
7141	AAGGCTTCAA	AAAGCACTCC	AGATGGCACA	ATAAAAGATC	GAAGATTGTT	TATGCATCAT
7201	GTTTCTTTAG	AGCCGATTAC	CTGTGTACCC	TTTCGCACAA	CTAAGGAACG	TCAAGAGATA
7261	CAGAATCCAA	ATTTTACCGC	ACCTGGTCAA	GAATTTCTGT	CTAAATCTCA	TTTGTATGAA
7321	CATCTGACTT	TGGAAAAATC	TTCAAGCAAT	TTAGCAGTTT	CAGGACATCC	ATTTTATCAA
7381	GTTTCTGCTA	CAAGAAATGA	AAAAATGAGA	CACTTGATTA	CTACAGGCAG	ACCAACCCAA
7441	GTCTTTGTTT	CACCTTTTAA	AACTAAATCA	CATTTTCACA	GAGTTGAACA	GTGTGTTAGG
7501	AATATTAAT	TGGAGGAAAA	CAGACAAAAG	CAAAACATTG	ATGGACATGG	CTCTGATGAT
7561	AGTAAAAATA	AGATTAATGA	CAATGAGATT	CATCAGTTTA	ACAAAAACAA	CTCCAATCAA
7621	GCAGCAGCTG	TAACTTTCAC	AAAGTGTGAA	GAAGAACCTT	TAGATTTAAT	TACAAGTCTT
7681	CAGAATGCCA	GAGATATACA	GGATATGCCA	ATTAAGAAGA	AACAAAGGCA	ACGCGTCTTT
7741	CCACAGCCAG	GCAGTCTGTA	TCTTGCAAAA	ACATCCACTC	TGCCTCGAAT	CTCTCTGAAA
7801	GCAGCAGTAG	GAGGCCAAGT	TCCCTCTGCG	TGTTCTCATA	AACAGCTGTA	TACGTATGGC
7861	GTTTCTAAAC	ATTGCATAAA	AATTAACAGC	AAAAATGCAG	AGTCTTTTCA	GTTTCACACT
7921	GAAGATTATT	TTGGTAAGGA	AAGTTTATGG	ACTGGAAAAG	GAATACAGTT	GGCTGATGGT
7981	GGATGGCTCA	TACCCTCCAA	TGATGGAAAAG	GCTGGAAAAG	AAGAATTTTA	TAGGGCTCTG
8041	TGTGACACTC	CAGGTGTGGA	TCCAAAGCTT	ATTTCTAGAA	TTTGGGTTTA	TAATCACTAT
8101	AGATGGATCA	TATGGAAACT	GGCAGCTATG	GAATGTGCCT	TTCTTAAGGA	ATTTGCTAAT
8161	AGATGCCTAA	GCCCAGAAAAG	GGTGCTTCTT	CAACTAAAAT	ACAGATATGA	TACGGAAATT
8221	GATAGAAGCA	GAAGATCGGC	TATAAAAAAG	ATAATGGAAA	GGGATGACAC	AGCTGCAAAA
8281	ACACTTGTTT	TCTGTGTTTC	TGACATAATT	TCATTGAGCG	CAATATATAT	TGAAACTTCT
8341	AGCAATAAAA	CTAGTAGTGC	AGATACCCAA	AAAGTGGCCA	TTATTGAACT	TACAGATGGG
8401	TGGTATGCTG	TTAAGGCCCA	GTTAGATCCT	CCCCTCTTAG	CTGTCTTAAA	GAATGGCAGA
8461	CTGACAGTTG	GTCAGAAGAT	TATTCTTCAT	GGAGCAGAAC	TGTTGGGCTC	TCCTGATGCC
8521	TGTACACCTC	TTGAAGCCCC	AGAATCTCTT	ATGTTAAAGA	TTTCTGCTAA	CAGTACTCGG
8581	CCTGCTCGCT	GGTATACCAA	ACTTGGAATC	TTTCTTGACC	CTAGACCTTT	TCCTCTGCCC
8641	TTATCATCGC	TTTTTCAGTGA	TGGAGGAAAT	GTTGGTTGTG	TTGATGTAAT	TATTCAAAGA
8701	GCATACCCTA	TACAGTGGAT	GGAGAAAGACA	TCATCTGGAT	TATACATATT	TCGCAATGAA
8761	AGAGAGGAAG	AAAAGGAAGC	AGCAAAATAT	GTGGAGGCCC	AACAAAAGAG	ACTAGAAGCC
8821	TTATTCACCTA	AAATTCAGGA	GGAATTTGAA	GAACATGAAG	AAAAACACAAC	AAAACCATAT
8881	TTACCATCAC	GTGCACTAAC	AAGACAGCAA	GTTCTGTGCT	TGCAAGATGG	TGCAGAGCTT
8941	TATGAAGCAG	TGAAGAATGC	AGCAGACCCA	GCTTACCTTG	AGGGTTATTT	CAGTGAAGAG
9001	CAGTTAAGAG	CCTTGAATAA	TCACAGGCCA	ATGTTGAATG	ATAAGAAAACA	AGCTCAGATC
9061	CAGTTGGAAG	TTAGGAAGGC	CATGGAATCT	GCTGAACAAA	AGGAACAAGG	TTTATCAAGG
9121	GATGTCACAA	CCGTGTGGAA	GTTGCGTATT	GTAAGCTATT	CAAAAAAGAA	AAAAGATTCA
9181	GTTATACTGA	GTATTTGGCG	TCCATCATCA	GATTTATATT	CTCTGTTAAC	AGAAGGAAAG
9241	AGATACAGAA	TTTATCATCT	TGCAACTTCA	AAATCTAAAA	GTAATCTCTGA	AAGAGCTAAC
9301	ATACAGTTAG	CAGCGACAAA	AAAAACTCAG	TATCAACAAC	TACCGGTTTC	AGATGAAATT
9361	TTATTTTCAGA	TTTACCAGCC	ACGGGAGCCC	CTTCACTTCA	GCAAATTTGT	AGATCCAGAC
9421	TTTCAGCCAT	TTTGTCTCTG	GCTGGACCTA	ATAGGATTTG	TCGTTTCTGT	TGTGAAAAAA
9481	ACAGGACTTG	CCCCTTTCGT	CTATTTGTCA	GACGAATGTT	ACAATTTTACT	GGCAATAAAG
9541	TTTGTGATAG	ACCTTAATGA	GGACATTATT	AAGCCTCATA	TGTTAATTGC	TGCAAGCAAC
9601	CTCCAGTGGC	GACCAGAATC	CAAATCAGGC	CTTCTTACTT	TATTTGCTGG	AGATTTTTCT
9661	GTGTTTTCTG	CTAGTCCAAA	AGAGGGCCAC	TTTCAAGAGA	CATTCAACAA	AATGAAAAAT
9721	ACTGTTGAGA	ATATTGACAT	ACTTTGCAAT	GAAGCAGAAA	ACAAGCTTAT	GCATATACTG

FIG. 3C

9781 CATGCAAATG ATCCCAAGTG GTCCACCCCA ACTAAAGACT GTACTTCAGG GCCGTACACT
9841 GCTCAAATCA TTCCTGGTAC AGGAAACAAG CTTCTGATGT CTTCTCCTAA TTGTGAGATA
9901 TATTATCAAA GTCCTTTATC ACTTTGTATG GCCAAAAGGA AGTCTGTTTC CACACCTGTC
9961 TCAGCCCAGA TGACTTCAAA GTCTTGTAAG GGGGAGAAAG AGATTGATGA CCAAAAAGAAC
10021 TGCAAAAAGA GAAGAGCCTT GGATTTCTTG AGTAGACTGC CTTTACCTCC ACCTGTTAGT
10081 CCCATTTGTA CATTTGTTTC TCCGGCTGCA CAGAAGGCAT TTCAGCCACC AAGGAGTTGT
10141 GGCACCAAAT ACGAAACACC CATAAAGAAA AAAGAACTGA ATTCTCCTCA GATGACTCCA
10201 TTTAAAAAAT TCAATGAAAT TTCTCTTTTG GAAAGTAATT CAATAGCTGA CGAAGAAGTT
10261 GCATTGATAA ATACCCAAGC TCTTTTGTCT GGTTCACAG GAGAAAAACA ATTTATATCT
10321 GTCAGTGAAT CCACTAGGAC TGCTCCACAC AGTTCAGAAG ATTATCTCAG ACTGAAACGA
10381 CGTTGTACTA CATCTCTGAT CAAAGAACAG GAGAGTTCCC AGGCCAGTAC GGAAGAATGT
10441 GAGAAAAATA AGCAGGACAC AATTACAAC AAAAATATA TCTAAGCATT TGCAAAGGCG
10501 ACAATAAATT ATTGACGCTT AACCTTTCCA GTTTATAAGA CTGGAATATA ATTTCAAACC
10561 ACACATTAGT ACTTATGTTG CACAATGAGA AAAGAAATTA GTTTCAAATT TACCTCAGCG
10621 TTTGTGTATC GGGCAAAAAT CGTTTGTCCC GATTCCGTAT TGGTATACTT TTGCTTCAGT
10681 TGCATATCTT AAAACTAAAT GTAATTTATT AACTAATCAA GAAAAACATC TTTGGCTGAG
10741 CTCGGTGGCT CATGCCTGTA ATCCCAACAC TTTGAGAAGC TGAGGTGGGA GGAGTGCTTG
10801 AGGCCAGGAG TTCAAGACCA GCCTGGGCAA CATAGGGAGA CCCCATCTT TACGAAGAAA
10861 AAAAAAAGG GGAAGAGAAA ATCTTTTAAA TCTTTGGATT TGATCACTAC AAGTATTATT
10921 TTACAAGTGA AATAAACATA CCATTTTCTT TTAGATTGTG TCATTAAATG GAATGAGGTC
10981 TCTTAGTACA GTTATTTTGA TGCAGATAAT TCCTTTTAGT TTAGCTACTA TTTTAGGGGA
11041 TTTTTTTTAG AGGTAACTCA CTATGAAATA GTTCTCCTTA ATGCAAATAT GTTGGTTCTG
11101 CTATAGTTCC ATCCTGTTC AAGTCAGGA TGAATATGAA GAGTGGTGT TTCTTTTGAG
11161 CAATTCTTCA TCCTTAAGTC AGCATGATTA TAAGAAAAAT AGAACCTCA GTGTAAGTCT
11221 AATTCCTTTT TACTATTCCA GTGTGATCTC TGAAATTAAA TTACTTCAAC TAAAAATTCA
11281 AATACTTTAA ATCAGAAGAT TTCATAGTTA ATTTATTTT TTTTCAACA AAATGGTCAT
11341 CAAACTCAA ACTTGAGAAA ATATCTTGCT TTCAAATTGA CACTA

FIG. 3D

Mar. 7, 2000

Sheet 7 of 9

6,033,857



A000455

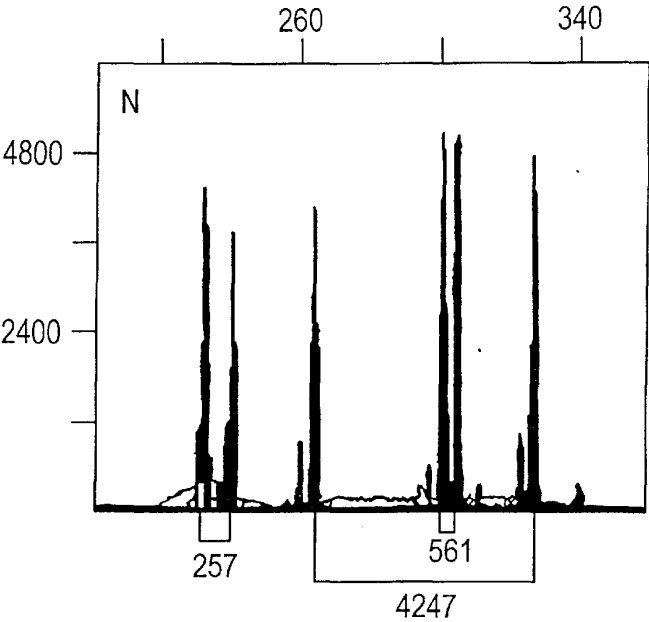


FIG. 5A

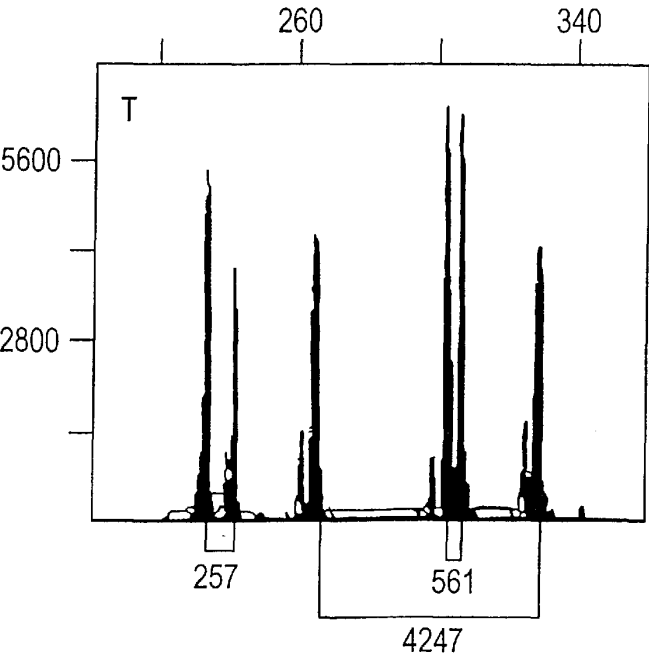


FIG. 5B

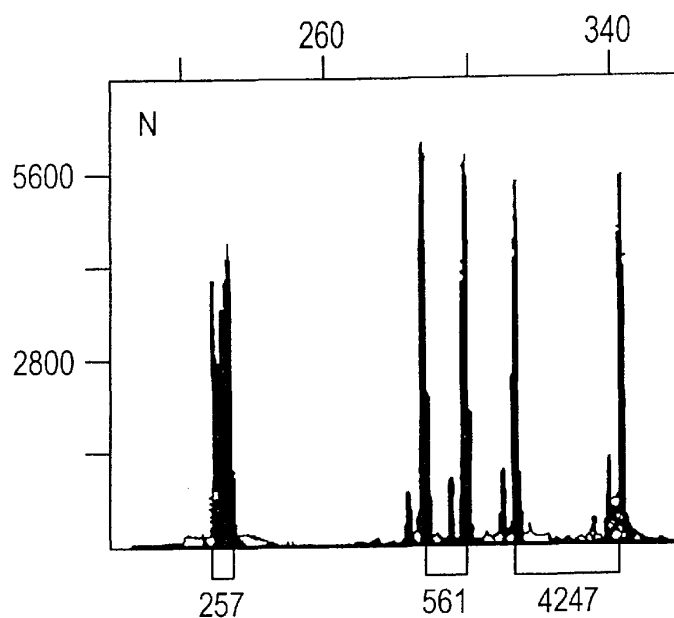


FIG. 5C

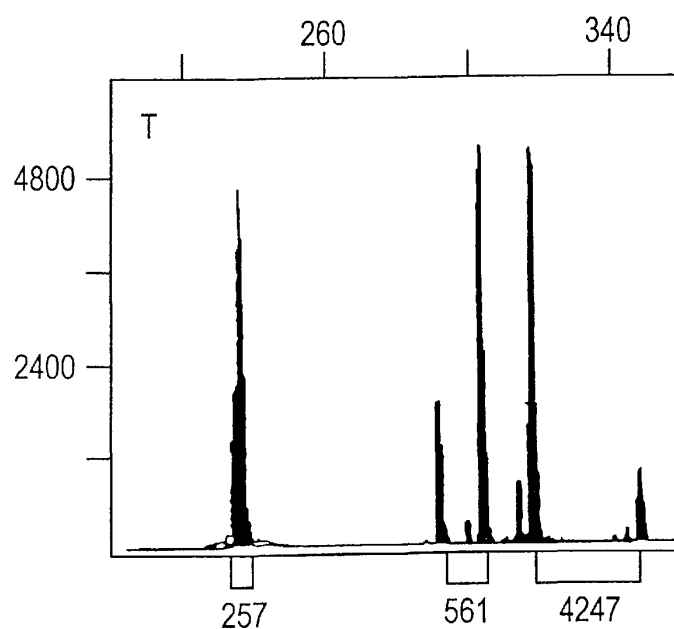


FIG. 5D

6,033,857

1

**CHROMOSOME 13-LINKED BREAST
CANCER SUSCEPTIBILITY GENE**

**CROSS REFERENCE TO RELATED
APPLICATION**

This application is a divisional of application Ser. No. 08/639,501, U.S. Pat. No. 5,837,492; filed on Apr. 29, 1996, U.S. Pat. No. 5,837,492; which is a continuation-in-part of application Ser. No. 08/585,391, filed on Jan. 11, 1996, now abandoned; which is a continuation-in-part of application Ser. No. 08/576,559 filed on Dec. 21, 1995, now abandoned; which is a continuation-in-part of application Ser. No. 08/575,359, filed on Dec. 20, 1995, now abandoned; which is a continuation-in-part of application Ser. No. 08/573,779, filed on Dec. 18, 1995, now abandoned; all of which are incorporated herein by reference.

FIELD OF THE INVENTION

The present invention relates generally to the field of human genetics. Specifically, the present invention relates to methods and materials used to isolate and detect a human cancer as predisposing gene (BRCA2), some mutant alleles of which cause susceptibility to cancer, in particular, breast cancer in females and males. More specifically, the invention relates to germline mutations in the BRCA2 gene and their use in the diagnosis of predisposition to breast cancer. The present invention further relates to somatic mutations in the BRCA2 gene in human breast cancer and their use in the diagnosis and prognosis of human breast cancer. Additionally, the invention relates to somatic mutations in the BRCA2 gene in other human cancers and their use in the diagnosis and prognosis of human cancers. The invention also relates to the therapy of human cancers which have a mutation in the BRCA2 gene, including gene therapy, protein replacement therapy and protein mimetics. The invention further relates to the screening of drugs for cancer therapy. Finally, the invention relates to the screening of the BRCA2 gene for mutations, which are useful for diagnosing the predisposition to breast cancer.

The publications and other materials used herein to illuminate the background of the invention, and in particular, cases to provide additional details respecting the practice, are incorporated herein by reference, and for convenience, are referenced by author and date in the following text and respectively grouped in the appended List of References.

BACKGROUND OF THE INVENTION

The genetics of cancer is complicated, involving multiple dominant, positive regulators of the transformed state (oncogenes) as well as multiple recessive, negative regulators (tumor suppressor genes). Over one hundred oncogenes have been characterized. Fewer than a dozen tumor suppressor genes have been identified, but the number is expected to increase beyond fifty (Knudson, 1993).

The involvement of so many genes underscores the complexity of the growth control mechanisms that operate in cells to maintain the integrity of normal tissue. This complexity is manifest in another way. So far, no single gene has been shown to participate in the development of all, or even the majority of human cancers. The most common oncogenic mutations are in the H-ras gene, found in 10–15% of all solid tumors (Anderson et al., 1992). The most frequently mutated tumor suppressor genes are the TP53 gene, homozygously deleted in roughly 50% of all tumors, and CDKN2, which was homozygously deleted in 46% of tumor

2

cell lines examined (Kamb et al., 1994a). Without a target that is common to all transformed cells, the dream of a “magic bullet” that can destroy or revert cancer cells while leaving normal tissue unharmed is improbable. The hope for a new generation of specifically targeted antitumor drugs may rest on the ability to identify tumor suppressor genes or oncogenes that play general roles in control of cell division.

The tumor suppressor genes which have been cloned and characterized influence susceptibility to: 1) Retinoblastoma (RB1); 2) Wilms’ tumor (WT1); 3) Li-Fraumeni (TP53); 4) Familial adenomatous polyposis (APC); 5) Neurofibromatosis type 1 (NF1); 6) Neurofibromatosis type 2 (NF2); 7) von Hippel-Lindau syndrome (VHL); 8) Multiple endocrine neoplasia type 2A (MEN2A), and 9) Melanoma (CDKN2).

Tumor suppressor loci that have been mapped genetically but not yet isolated include genes for: Multiple endocrine neoplasia type 1 (MEN1); Lynch cancer family syndrome 2 (LCFS2); Neuroblastoma (NB); Basal cell nevus syndrome (BCNS); Beckwith-Wiedemann syndrome (BWS); Renal cell carcinoma (RCC); Tuberous sclerosis 1 (TSC1); and Tuberous sclerosis 2 (TSC2). The tumor suppressor genes that have been characterized to date encode products with similarities to a variety of protein types, including DNA binding proteins (WT1), ancillary transcription regulators (RB1), GTPase activating proteins or GAPs (NF1), cytoskeletal components (NF2), membrane bound receptor kinases (MEN2A), cell cycle regulators (CDKN2) and others with no obvious similarity to known proteins (APC and VHL).

In many cases, the tumor suppressor gene originally identified through genetic studies has been shown to be lost or mutated in some sporadic tumors. This result suggests that regions of chromosomal aberration may signify the position of important tumor suppressor genes involved both in genetic predisposition to cancer and in sporadic cancer.

One of the hallmarks of several tumor suppressor genes characterized to date is that they are deleted at high frequency in certain tumor types. The deletions often involve loss of a single allele, a so-called loss of heterozygosity (LOH), but may also involve homozygous deletion of both alleles. For LOH, the remaining allele is presumed to be nonfunctional, either because of a preexisting inherited mutation, or because of a secondary sporadic mutation. Breast cancer is one of the most significant diseases that affects women. At the current rate, American women have a 1 in 8 risk of developing breast cancer by age 95 (American Cancer Society, 1992). Treatment of breast cancer at later stages is often futile and disfiguring, making early detection a high priority in medical management of the disease. Ovarian cancer, although less frequent than breast cancer, is often rapidly fatal and is the fourth most common cause of cancer mortality in American women. Genetic factors contribute to an ill-defined proportion of breast cancer incidence, estimated to be about 5% of all cases but approximately 25% of cases diagnosed before age 40 (Claus et al., 1991). Breast cancer has been subdivided into two types, early-age onset and late-age onset, based on an inflection in the age-specific incidence curve around age 50. Mutation of one gene, BRCA1, is thought to account for approximately 45% of familial breast cancer, but at least 80% of families with both breast and ovarian cancer (Easton et al., 1993).

The BRCA1 gene has been isolated (Futreal et al., 1994; Miki et al., 1994) following an intense effort following its mapping in 1990 (Hall et al., 1990; Narod et al., 1991). A second locus, BRCA2, has recently been mapped to chromosome 13 (Wooster et al., 1994) and appears to account for a proportion of early-onset breast cancer roughly equal to

BRCA1, but confers a lower risk of ovarian cancer. The remaining susceptibility to early-onset breast cancer is divided between as-yet unmapped genes for familial cancer, and rarer germline mutations in genes such as TP53 (Malkin et al., 1990). It has also been suggested that heterozygote carriers for defective forms of the Ataxia-Telangiectasia gene are at higher risk for breast cancer (Swift et al., 1976; Swift et al., 1991). Late-age onset breast cancer is also often familial although the risks in relatives are not as high as those for early-onset breast cancer (Cannon-Albright et al., 1994; Mettlin et al., 1990). However, the percentage of such cases due to genetic susceptibility is unknown.

Breast cancer has long been recognized to be, in part, a familial disease (Anderson, 1972). Numerous investigators have examined the evidence for genetic inheritance and concluded that the data are most consistent with dominant inheritance for a major susceptibility locus or loci (Bishop and Gardner, 1980; Go et al., 1983; Williams and Anderson, 1984; Bishop et al., 1988; Newman et al., 1988; Claus et al., 1991). Recent results demonstrate that at least three loci exist which convey susceptibility to breast cancer as well as other cancers. These loci are the TP53 locus on chromosome 17p (Malkin et al., 1990), a 17q-linked susceptibility locus known as BRCA1 (Hall et al., 1990), and one or more loci responsible for the unmapped residual. Hall et al. (1990) indicated that the inherited breast cancer susceptibility in kindreds with early age onset is linked to chromosome 17q21; although subsequent studies by this group using a more appropriate genetic model partially refuted the limitation to early onset breast cancer (Margaritte et al., 1992).

Most strategies for cloning the chromosome 13-linked breast cancer predisposing gene (BRCA2) require precise genetic localization studies. The simplest model for the functional role of BRCA2 holds that alleles of BRCA2 that predispose to cancer are recessive to wild type alleles; that is, cells that contain at least one wild type BRCA2 allele are not cancerous. However, cells that contain one wild type BRCA2 allele and one predisposing allele may occasionally suffer loss of the wild type allele either by random mutation or by chromosome loss during cell division (nondisjunction). All the progeny of such a mutant cell lack the wild type function of BRCA2 and may develop into tumors. According to this model, predisposing alleles of BRCA2 are recessive, yet susceptibility to cancer is inherited in a dominant fashion: women who possess one predisposing allele (and one wild type allele) risk developing cancer, because their mammary epithelial cells may spontaneously lose the wild type BRCA2 allele. This model applies to a group of cancer susceptibility loci known as tumor suppressors or antioncogenes, a class of genes that includes the retinoblastoma gene and neurofibromatosis gene. By inference this model may explain the BRCA1 function, as has recently been suggested (Smith et al., 1992).

A second possibility is that BRCA2 predisposing alleles are truly dominant; that is, a wild type allele of BRCA2 cannot overcome the tumor forming role of the predisposing allele. Thus, a cell that carries both wild type and mutant alleles would not necessarily lose the wild type copy of BRCA2 before giving rise to malignant cells. Instead, mammary cells in predisposed individuals would undergo some other stochastic change(s) leading to cancer.

If BRCA2 predisposing alleles are recessive, the BRCA2 gene is expected to be expressed in normal mammary tissue but not functionally expressed in mammary tumors. In contrast, if BRCA2 predisposing alleles are dominant, the wild type BRCA2 gene may or may not be expressed in normal mammary tissue. However, the predisposing allele will likely be expressed in breast tumor cells.

The chromosome 13 linkage of BRCA2 was independently confirmed by studying fifteen families that had multiple cases of early-onset breast cancer cases that were not linked to BRCA1 (Wooster et al., 1994). These studies claimed to localize the gene within a large region, 6 centi-Morgans (cM), or approximately 6 million base pairs, between the markers D13S289 and D13S267, placing BRCA2 in a physical region defined by 13q12-13. The size of these regions and the uncertainty associated with them has made it difficult to design and implement physical mapping and/or cloning strategies for isolating the BRCA2 gene. Like BRCA1, BRCA2 appears to confer a high risk of early-onset breast cancer in females. However, BRCA2 does not appear to confer a substantially elevated risk of ovarian cancer, although it does appear to confer an elevated risk of male breast cancer (Wooster, et al., 1994).

Identification of a breast cancer susceptibility locus would permit the early detection of susceptible individuals and greatly increase our ability to understand the initial steps which lead to cancer. As susceptibility loci are often altered during tumor progression, cloning these genes could also be important in the development of better diagnostic and prognostic products, as well as better cancer therapies.

SUMMARY OF THE INVENTION

The present invention relates generally to the field of human genetics. Specifically, the present invention relates to methods and materials used to isolate and detect a human breast cancer predisposing gene (BRCA2), some alleles of which cause susceptibility to cancer, in particular breast cancer in females and males. More specifically, the present invention relates to germline mutations in the BRCA2 gene and their use in the diagnosis of predisposition to breast cancer. The invention further relates to somatic mutations in the BRCA2 gene in human breast cancer and their use in the diagnosis and prognosis of human breast cancer. Additionally, the invention relates to somatic mutations in the BRCA2 gene in other human cancers and their use in the diagnosis and prognosis of human cancers. The invention also relates to the therapy of human cancers which have a mutation in the BRCA2 gene, including gene therapy, protein replacement therapy and protein mimetics. The invention further relates to the screening of drugs for cancer therapy. Finally, the invention relates to the screening of the BRCA2 gene for mutations, which are useful for diagnosing the predisposition to breast cancer.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 shows a schematic map of STSs, PIs, BACs and YACs in the BRCA2 region.

FIG. 2 shows the sequence-space relationship between the cDNA clones, hybrid selected clones, cDNA PCR products and genomic sequences used to assemble the BRCA2 transcript sequence. 2-Br-C:RACE is a biotin-capture RACE product obtained from both human breast and human thymus cDNA. The cDNA clone λ sc713.1 was identified by screening a pool of human testis and HepG2 cDNA libraries with hybrid selected clone GT 713. The sequence 1-BR:CG026 \rightarrow 5 kb was generated from a PCR product beginning at the exon 7 junction (within λ sc713.1) and terminating within an hybrid selected clone that is part of exon 11. The sequence of exon 11 was corrected by comparison to hybrid selected clones, genomic sequence in the public domain and radioactive DNA sequencing gels. Hybrid selected clones located within that exon (clone names beginning with nH or GT) are placed below it. The

cDNA clones λ wCBF1B8.1, λ wCBF1A5.1, λ wCBF1A5.12, λ wCBF1B6.2 and λ wCBF1B6.3 were identified by screening a pool of human mammary gland, placenta, testis and HepG2 cDNA libraries with the exon trapped clones wXBF1B8, wXPF1A5 and wXBF1B6. The clone λ wCBF1B6.3 is chimeric (indicated by the dashed line), but its 5' end contained an important overlap with λ wCBF1A5.1. denotes the translation initiator. denotes the translation terminator.

FIGS. 3A–3D show the DNA sequence of the BRCA2 gene (which is also set forth in SEQ ID NO:1).

FIG. 4 shows the genomic organization of the BRCA2 gene. The exons (boxes and/or vertical lines) are parsed across the genomic sequences (ftp://genome.wustl.edu/pub/gsc/brca;) (horizontal lines) such that their sizes and spacing are proportional. The name of each genomic sequence is given at the left side of the figure. The sequences 92M18.00541 and 92M18.01289 actually overlap. Distances between the other genomic sequences are not known. Neither the public database nor our sequence database contained genomic sequences overlapping with exon 21.

Exons 1, 11 and 21 are numbered. “*” denotes two adjacent exons spaced closely enough that they are not resolved at this scale.

FIGS. 5A–5D show a loss of heterozygosity (LOH) analysis of primary breast tumors. Alleles of STR markers are indicated below the chromatogram. Shown are one example of a tumor heterozygous at BRCA2 (FIGS. 5A and 5B) and an example of a tumor with LOH at BRCA2 (FIGS. 5C and 5D). Fluorescence units are on the ordinate; size in basepairs is on the abscissa. N is for normal (FIGS. 5A and 5C) and T is for tumor (FIGS. 5B and 5D).

DETAILED DESCRIPTION OF THE INVENTION

The present invention relates generally to the field of human genetics. Specifically, the present invention relates to methods and materials used to isolate and detect a human breast cancer predisposing gene (BRCA2), some alleles of which cause susceptibility to cancer, in particular breast cancer in females and males. More specifically, the present invention relates to germline mutations in the BRCA2 gene and their use in the diagnosis of predisposition to breast cancer. The invention further relates to somatic mutations in the BRCA2 gene in human breast cancer and their use in the diagnosis and prognosis of human breast cancer. Additionally, the invention relates to somatic mutations in the BRCA2 gene in other human cancers and their use in the diagnosis and prognosis of human cancers. The invention also relates to the therapy of human cancers which have a mutation in the BRCA2 gene, including gene therapy, protein replacement therapy and protein mimetics. The invention further relates to the screening of drugs for cancer therapy. Finally, the invention relates to the screening of the BRCA2 gene for mutations, which are useful for diagnosing the predisposition to breast cancer.

The present invention provides an isolated polynucleotide comprising all, or a portion of the BRCA2 locus or of a mutated BRCA2 locus, preferably at least eight bases and not more than about 100 kb in length. Such polynucleotides may be antisense polynucleotides. The present invention also provides a recombinant construct comprising such an isolated polynucleotide, for example, a recombinant construct suitable for expression in a transformed host cell.

Also provided by the present invention are methods of detecting a polynucleotide comprising a portion of the

BRCA2 locus or its expression product in an analyte. Such methods may further comprise the step of amplifying the portion of the BRCA2 locus, and may further include a step of providing a set of polynucleotides which are primers for amplification of said portion of the BRCA2 locus. The method is useful for either diagnosis of the predisposition to cancer or the diagnosis or prognosis of cancer.

The present invention also provides isolated antibodies, preferably monoclonal antibodies, which specifically bind to an isolated polypeptide comprised of at least five amino acid residues encoded by the BRCA2 locus.

The present invention also provides kits for detecting in an analyte a polynucleotide comprising a portion of the BRCA2 locus, the kits comprising a polynucleotide complementary to the portion of the BRCA2 locus packaged in a suitable container, and instructions for its use.

The present invention further provides methods of preparing a polynucleotide comprising polymerizing nucleotides to yield a sequence comprised of at least eight consecutive nucleotides of the BRCA2 locus; and methods of preparing a polypeptide comprising polymerizing amino acids to yield a sequence comprising at least five amino acids encoded within the BRCA2 locus.

The present invention further provides methods of screening the BRCA2 gene to identify mutations. Such methods may further comprise the step of amplifying a portion of the BRCA2 locus, and may further include a step of providing a set of polynucleotides which are primers for amplification of said portion of the BRCA2 locus. The method is useful for identifying mutations for use in either diagnosis of the predisposition to cancer or the diagnosis or prognosis of cancer.

The present invention further provides methods of screening suspected BRCA2 mutant alleles to identify mutations in the BRCA2 gene.

In addition, the present invention provides methods of screening drugs for cancer therapy to identify suitable drugs for restoring BRCA2 gene product function.

Finally, the present invention provides the means necessary for production of gene-based therapies directed at cancer cells. These therapeutic agents may take the form of polynucleotides comprising all or a portion of the BRCA2 locus placed in appropriate vectors or delivered to target cells in more direct ways such that the function of the BRCA2 protein is reconstituted. Therapeutic agents may also take the form of polypeptides based on either a portion of, or the entire protein sequence of BRCA2. These may functionally replace the activity of BRCA2 in vivo.

It is a discovery of the present invention that the BRCA2 locus which predisposes individuals to breast cancer, is a gene encoding a BRCA2 protein. This gene is termed BRCA2 herein. It is a discovery of the present invention that mutations in the BRCA2 locus in the germline are indicative of a predisposition to breast cancer in both men and women. Finally, it is a discovery of the present invention that somatic mutations in the BRCA2 locus are also associated with breast cancer and other cancers, which represents an indicator of these cancers or of the prognosis of these cancers. The mutational events of the BRCA2 locus can involve deletions, insertions and point mutations within the coding sequence and the non-coding sequence.

Starting from a region on human chromosome 13 of the human genome, which has a size estimated at about 6 million base pairs, a smaller region of 1 to 1.5 million bases which contains a genetic locus, BRCA2, which causes susceptibility to cancer, including breast cancer, has been identified.

6,033,857

7

The region containing the BRCA2 locus was identified using a variety of genetic techniques. Genetic mapping techniques initially defined the BRCA2 region in terms of recombination with genetic markers. Based upon studies of large extended families ("kindreds") with multiple cases of breast cancer, a chromosomal region has been pinpointed that contains the BRCA2 gene. A region which contains the BRCA2 locus is physically bounded by the markers D13S289 and D13S267.

The use of the genetic markers provided by this invention allowed the identification of clones which cover the region from a human yeast artificial chromosome (YAC) or a human bacterial artificial chromosome (BAC) library. It also allowed for the identification and preparation of more easily manipulated P1 and BAC clones from this region and the construction of a contig from a subset of the clones. These P1s, YACs and BACs provide the basis for cloning the BRCA2 locus and provide the basis for developing reagents effective, for example, in the diagnosis and treatment of breast and/or ovarian cancer. The BRCA2 gene and other potential susceptibility genes have been isolated from this region. The isolation was done using software trapping (a computational method for identifying sequences likely to contain coding exons, from contiguous or discontinuous genomic DNA sequences), hybrid selection techniques and direct screening, with whole or partial cDNA inserts from P1s and BACs, in the region to screen cDNA libraries. These methods were used to obtain sequences of loci expressed in breast and other tissue. These candidate loci were analyzed to identify sequences which confer cancer susceptibility. We have discovered that there are mutations in the coding sequence of the BRCA2 locus in kindreds which are responsible for the chromosome 13-linked cancer susceptibility known as BRCA2. The present invention not only facilitates the early detection of certain cancers, so vital to patient survival, but also permits the detection of susceptible individuals before they develop cancer.

Population Resources

Large, well-documented Utah kindreds are especially important in providing good resources for human genetic studies. Each large kindred independently provides the power to detect whether a BRCA2 susceptibility allele is segregating in that family. Recombinants informative for localization and isolation of the BRCA2 locus could be obtained only from kindreds large enough to confirm the presence of a susceptibility allele. Large sibships are especially important for studying breast cancer, since penetrance of the BRCA2 susceptibility allele is reduced both by age and sex, making informative sibships difficult to find. Furthermore, large sibships are essential for constructing haplotypes of deceased individuals by inference from the haplotypes of their close relatives.

While other populations may also provide beneficial information, such studies generally require much greater effort, and the families are usually much smaller and thus less informative. Utah's age-adjusted breast cancer incidence is 20% lower than the average U.S. rate. The lower incidence in Utah is probably due largely to an early age at first pregnancy, increasing the probability that cases found in Utah kindreds carry a genetic predisposition.

Genetic Mapping

Given a set of informative families, genetic markers are essential for linking a disease to a region of a chromosome. Such markers include restriction fragment length polymorphisms (RFLPs) (Botstein et al., 1980), markers with a variable number of tandem repeats (VNTRs) (Jeffreys et al., 1985, Nakamura et al., 1987), and an abundant class of DNA

8

polymorphisms based on short tandem repeats (STRs), especially repeats of CpA (Weber and May, 1989; Litt et al., 1989). To generate a genetic map, one selects potential genetic markers and tests them using DNA extracted from members of the kindreds being studied.

Genetic markers useful in searching for a genetic locus associated with a disease can be selected on an ad hoc basis, by densely covering a specific chromosome, or by detailed analysis of a specific region of a chromosome. A preferred method for selecting genetic markers linked with a disease involves evaluating the degree of informativeness of kindreds to determine the ideal distance between genetic markers of a given degree of polymorphism, then selecting markers from known genetic maps which are ideally spaced for maximal efficiency. Informativeness of kindreds is measured by the probability that the markers will be heterozygous in unrelated individuals. It is also most efficient to use STR markers which are detected by amplification of the target nucleic acid sequence using PCR; such markers are highly informative, easy to assay (Weber and May, 1989), and can be assayed simultaneously using multiplexing strategies (Skolnick and Wallace, 1988), greatly reducing the number of experiments required.

Once linkage has been established, one needs to find markers that flank the disease locus, i.e., one or more markers proximal to the disease locus, and one or more markers distal to the disease locus. Where possible, candidate markers can be selected from a known genetic map. Where none is known, new markers can be identified by the STR technique, as shown in the Examples.

Genetic mapping is usually an iterative process. In the present invention, it began by defining flanking genetic markers around the BRCA2 locus, then replacing these flanking markers with other markers that were successively closer to the BRCA2 locus. As an initial step, recombination events, defined by large extended kindreds, helped specifically to localize the BRCA2 locus as either distal or proximal to a specific genetic marker (Wooster et al., 1994).

The region surrounding BRCA2, until the disclosure of the present invention, was not well mapped and there were few markers. Therefore, short repetitive sequences were developed from cosmids, P1s, BACs and YACs, which physically map to the region and were analyzed in order to develop new genetic markers. Novel STRs were found which were both polymorphic and which mapped to the BRCA2 region.

Physical Mapping

Three distinct methods were employed to physically map the region. The first was the use of yeast artificial chromosomes (YACs) to clone the BRCA2 region. The second was the creation of a set of P1, BAC and cosmid clones which cover the region containing the BRCA2 locus.

Yeast Artificial Chromosomes (YACs). Once a sufficiently small region containing the BRCA2 locus was identified, physical isolation of the DNA in the region proceeded by identifying a set of overlapping YACs which covers the region. Useful YACs can be isolated from known libraries, such as the St. Louis and CEPH YAC libraries, which are widely distributed and contain approximately 50,000 YACs each. The YACs isolated were from these publicly accessible libraries and can be obtained from a number of sources including the Michigan Genome Center.

Clearly, others who had access to these YACs, without the disclosure of the present invention, would not have known the value of the specific YACs we selected since they would not have known which YACs were within, and which YACs outside of, the smallest region containing the BRCA2 locus.

6,033,857

9

P1 and BAC Clones. In the present invention, it is advantageous to proceed by obtaining P1 and BAC clones to cover this region. The smaller size of these inserts, compared to YAC inserts, makes them more useful as specific hybridization probes. Furthermore, having the cloned DNA in bacterial cells, rather than in yeast cells, greatly increases the ease with which the DNA of interest can be manipulated, and improves the signal-to-noise ratio of hybridization assays.

P1 and BAC clones are obtained by screening libraries constructed from the total human genome with specific sequence tagged sites (STSS) derived from the YACs, P1s and BACs, isolated as described herein.

These P1 and BAC clones can be compared by interspersed repetitive sequence (IRS) PCR and/or restriction enzyme digests followed by gel electrophoresis and comparison of the resulting DNA fragments ("fingerprints") (Maniatis et al., 1982). The clones can also be characterized by the presence of STSS. The fingerprints are used to define an overlapping contiguous set of clones which covers the region but is not excessively redundant, referred to herein as a "minimum tiling path". Such a minimum tiling path forms the basis for subsequent experiments to identify cDNAs which may originate from the BRCA2 locus.

P1 clones (Sternberg, 1990; Sternberg et al., 1990; Pierce et al., 1992; Shizuya et al., 1992) were isolated by Genome Sciences using PCR primers provided by us for screening. BACs were provided by hybridization techniques in Dr. Mel Simon's laboratory and by analysis of PCR pools in our laboratory. The strategy of using P1 and BAC clones also permitted the covering of the genomic region with an independent set of clones not derived from YACs. This guards against the possibility of deletions in YACs. These new sequences derived from the P1 and BAC clones provide the material for further screening for candidate genes, as described below.

Gene Isolation.

There are many techniques for testing genomic clones for the presence of sequences likely to be candidates for the coding sequence of a locus one is attempting to isolate, including but not limited to: (a) zoo blots, (b) identifying HTF islands, (c) exon trapping, (d) hybridizing cDNA to P1s, BAC or YACs and (e) screening cDNA libraries.

(a) Zoo blots. The first technique is to hybridize cosmids to Southern blots to identify DNA sequences which are evolutionarily conserved, and which therefore give positive hybridization signals with DNA from species of varying degrees of relationship to humans (such as monkey, cow, chicken, pig, mouse and rat). Southern blots containing such DNA from a variety of species are commercially available (Clontech, Cat. 7753-1).

(b) Identifying HTF islands. The second technique involves finding regions rich in the nucleotides C and G, which often occur near or within coding sequences. Such sequences are called HTF (HpaI tiny fragment) or CpG islands, as restriction enzymes specific for sites which contain CpG dimers cut frequently in these regions (Lindsay et al., 1987).

(c) Exon trapping. The third technique is exon trapping, a method that identifies sequences in genomic DNA which contain splice junctions and therefore are likely to comprise coding sequences of genes. Exon amplification (Buckler et al., 1991) is used to select and amplify exons from DNA clones described above. Exon amplification is based on the selection of RNA sequences which are flanked by functional 5' and/or 3' splice sites. The products of the exon amplification are

10

used to screen the breast cDNA libraries to identify a manageable number of candidate genes for further study. Exon trapping can also be performed on small segments of sequenced DNA using computer programs or by software trapping.

(d) Hybridizing cDNA to P1s, BACs or YACs. The fourth technique is a modification of the selective enrichment technique which utilizes hybridization of cDNA to cosmids, P1s, BACs or YACs and permits transcribed sequences to be identified in, and recovered from cloned genomic DNA (Kandpal et al., 1990). The selective enrichment technique, as modified for the present purpose, involves binding DNA from the region of BRCA2 present in a YAC to a column matrix and selecting cDNAs from the relevant libraries which hybridize with the bound DNA, followed by amplification and purification of the bound DNA, resulting in a great enrichment for cDNAs in the region represented by the cloned genomic DNA.

(e) Identification of cDNAs. The fifth technique is to identify cDNAs that correspond to the BRCA2 locus. Hybridization probes containing putative coding sequences, selected using any of the above techniques, are used to screen various libraries, including breast tissue cDNA libraries and any other necessary libraries.

Another variation on the theme of direct selection of cDNA can be used to find candidate genes for BRCA2 (Lovett et al., 1991; Futreal, 1993). This method uses cosmid, P1 or BAC DNA as the probe. The probe DNA is digested with a blunt cutting restriction enzyme such as HaeIII. Double stranded adapters are then ligated onto the DNA and serve as binding sites for primers in subsequent PCR amplification reactions using biotinylated primers. Target cDNA is generated from mRNA derived from tissue samples, e.g., breast tissue, by synthesis of either random primed or oligo(dT) primed first strand followed by second strand synthesis. The cDNA ends are rendered blunt and ligated onto double-stranded adapters. These adapters serve as amplification sites for PCR. The target and probe sequences are denatured and mixed with human C₀t-1 DNA to block repetitive sequences. Solution hybridization is carried out to high C₀t-1/2 values to ensure hybridization of rare target cDNA molecules. The annealed material is then captured on avidin beads, washed at high stringency and the retained cDNAs are eluted and amplified by PCR. The selected cDNA is subjected to further rounds of enrichment before cloning into a plasmid vector for analysis.

Testing the cDNA for Candidacy

Proof that the cDNA is the BRCA2 locus is obtained by finding sequences in DNA extracted from affected kindred members which create abnormal BRCA2 gene products or abnormal levels of BRCA2 gene product. Such BRCA2 susceptibility alleles will co-segregate with the disease in large kindreds. They will also be present at a much higher frequency in non-kindred individuals with breast cancer than in individuals in the general population. Finally, since tumors often mutate somatically at loci which are in other instances mutated in the germline, we expect to see normal germline BRCA2 alleles mutated into sequences which are identical or similar to BRCA2 susceptibility alleles in DNA extracted from tumor tissue. Whether one is comparing BRCA2 sequences from tumor tissue to BRCA2 alleles from the germline of the same individuals, or one is comparing germline BRCA2 alleles from cancer cases to those from unaffected individuals, the key is to find mutations which are serious enough to cause obvious disruption to the normal function of the gene product. These mutations can take a

6,033,857

11

number of forms. The most severe forms would be frame shift mutations or large deletions which would cause the gene to code for an abnormal protein or one which would significantly alter protein expression. Less severe disruptive mutations would include small in-frame deletions and non-conservative base pair substitutions which would have a significant effect on the protein produced, such as changes to or from a cysteine residue, from a basic to an acidic amino acid or vice versa, from a hydrophobic to hydrophilic amino acid or vice versa, or other mutations which would affect secondary, tertiary or quaternary protein structure. Silent mutations or those resulting in conservative amino acid substitutions would not generally be expected to disrupt protein function.

According to the diagnostic and prognostic method of the present invention, alteration of the wild-type BRCA2 locus is detected. In addition, the method can be performed by detecting the wild-type BRCA2 locus and confirming the lack of a predisposition to cancer at the BRCA2 locus. "Alteration of a wild-type gene" encompasses all forms of mutations including deletions, insertions and point mutations in the coding and noncoding regions. Deletions may be of the entire gene or of only a portion of the gene. Point mutations may result in stop codons, frameshift mutations or amino acid substitutions. Somatic mutations are those which occur only in certain tissues, e.g., in the tumor tissue, and are not inherited in the germline. Germline mutations can be found in any of a body's tissues and are inherited. If only a single allele is somatically mutated, an early neoplastic state is indicated. However, if both alleles are somatically mutated, then a late neoplastic state is indicated. The finding of BRCA2 mutations thus provides both diagnostic and prognostic information. A BRCA2 allele which is not deleted (e.g., found on the sister chromosome to a chromosome carrying a BRCA2 deletion) can be screened for other mutations, such as insertions, small deletions, and point mutations. It is believed that many mutations found in tumor tissues will be those leading to decreased expression of the BRCA2 gene product. However, mutations leading to non-functional gene products would also lead to a cancerous state. Point mutational events may occur in regulatory regions. Such as in the promoter of the gene, leading to loss or diminution of expression of the mRNA. Point mutations may also abolish proper RNA processing, leading to loss of expression of the BRCA2 gene product, or to a decrease in mRNA stability or translation efficiency.

Useful diagnostic techniques include, but are not limited to fluorescent in situ hybridization (FISH), direct DNA sequencing, PFGE analysis, Southern blot analysis, single stranded conformation analysis (SSCA), RNase protection assay, allele-specific oligonucleotide (ASO), dot blot analysis and PCR-SSCP, as discussed in detail further below.

Predisposition to cancers, such as breast cancer, and the other cancers identified herein, can be ascertained by testing any tissue of a human for mutations of the BRCA2 gene. For example, a person who has inherited a germline BRCA2 mutation would be prone to develop cancers. This can be determined by testing DNA from any tissue of the person's body. Most simply, blood can be drawn and DNA extracted from the cells of the blood. In addition, prenatal diagnosis can be accomplished by testing fetal cells, placental cells or amniotic cells for mutations of the BRCA2 gene. Alteration of a wild-type BRCA2 allele, whether, for example, by point mutation or deletion, can be detected by any of the means discussed herein.

There are several methods that can be used to detect DNA sequence variation. Direct DNA sequencing, either manual

12

sequencing or automated fluorescent sequencing can detect sequence variation. For a gene as large as BRCA2, manual sequencing is very labor-intensive, but under optimal conditions, mutations in the coding sequence of a gene are rarely missed. Another approach is the single-stranded conformation polymorphism assay (SSCA) (Orita et al., 1989). This method does not detect all sequence changes, especially if the DNA fragment size is greater than 200 bp, but can be optimized to detect most DNA sequence variation. The reduced detection sensitivity is a disadvantage but the increased throughput possible with SSCA makes it an attractive, viable alternative to direct sequencing for mutation detection on a research basis. The fragments which have shifted mobility on SSCA gels are then sequenced to determine the exact nature of the DNA sequence variation. Other approaches based on the detection of mismatches between the two complementary DNA strands include clamped denaturing gel electrophoresis (CDGE) (Sheffield et al., 1991), heteroduplex analysis (HA) (White et al., 1992) and chemical mismatch cleavage (CMC) (Grompe et al., 1989). None of the methods described above will detect large deletions, duplications or insertions, nor will they detect a regulatory mutation which affects transcription or translation of the protein. Other methods which might detect these classes of mutations such as a protein truncation assay or the asymmetric assay, detect only specific types of mutations and would not detect missense mutations. A review of currently available methods of detecting DNA sequence variation can be found in a recent review by Grompe (1993). Once a mutation is known, an allele specific detection approach such as allele specific oligonucleotide (ASO) hybridization can be utilized to rapidly screen large numbers of other samples for that same mutation.

In order to detect the alteration of the wild-type BRCA2 gene in a tissue, it is helpful to isolate the tissue free from surrounding normal tissues. Means for enriching tissue preparation for tumor cells are known in the art. For example, the tissue may be isolated from paraffin or cryostat sections. Cancer cells may also be separated from normal cells by flow cytometry. These techniques, as well as other techniques for separating tumor cells from normal cells, are well known in the art. If the tumor tissue is highly contaminated with normal cells, detection of mutations is more difficult.

A rapid preliminary analysis to detect polymorphisms in DNA sequences can be performed by looking at a series of Southern blots of DNA cut with one or more restriction enzymes, preferably with a large number of restriction enzymes. Each blot contains a series of normal individuals and a series of cancer cases, tumors, or both. Southern blots displaying hybridizing fragments (differing in length from control DNA when probed with sequences near or including the BRCA2 locus) indicate a possible mutation. If restriction enzymes which produce very large restriction fragments are used, then pulsed field gel electrophoresis (PFGE) is employed.

Detection of point mutations may be accomplished by molecular cloning of the BRCA2 allele(s) and sequencing the allele(s) using techniques well known in the art. Alternatively, the gene sequences can be amplified directly from a genomic DNA preparation from the tumor tissue, using known techniques. The DNA sequence of the amplified sequences can then be determined.

There are six well known methods for a more complete, yet still indirect, test for confirming the presence of a susceptibility allele: 1) single stranded conformation analysis (SSCA) (Orita et al., 1989); 2) denaturing gradient gel

electrophoresis (DGGE) (Wartell et al., 1990; Sheffield et al., 1989); 3) RNase protection assays (Finkelstein et al., 1990; Kinszler et al., 1991); 4) allele-specific oligonucleotides (ASOs) (Conner et al., 1983); 5) the use of proteins which recognize nucleotide mismatches, such as the *E. coli* mutS protein (Modrich, 1991); and 6) allele-specific PCR (Rano & Kidd, 1989). For allele-specific PCR, primers are used which hybridize at their 3' ends to a particular BRCA2 mutation. If the particular BRCA2 mutation is not present, an amplification product is not observed. Amplification Refractory Mutation System (ARMS) can also be used, as disclosed in European Patent Application Publication No. 0332435 and in Newton et al., 1989.

Insertions and deletions of genes can also be detected by cloning, sequencing and amplification. In addition, restriction fragment length polymorphism (RFLP) probes for the gene or surrounding marker genes can be used to score alteration of an allele or an insertion in a polymorphic fragment. Such a method is particularly useful for screening relatives of an affected individual for the presence of the BRCA2 mutation found in that individual. Other techniques for detecting insertions and deletions as known in the art can be used.

In the first three methods (SSCA, DGGE and RNase protection assay), a new electrophoretic band appears. SSCA detects a band which migrates differentially because the sequence change causes a difference in single-strand, intramolecular base pairing. RNase protection involves cleavage of the mutant polynucleotide into two or more smaller fragments. DGGE detects differences in migration rates of mutant sequences compared to wild-type sequences, using a denaturing gradient gel. In an allele-specific oligonucleotide assay, an oligonucleotide is designed which detects a specific sequence, and the assay is performed by detecting the presence or absence of a hybridization signal. In the mutS assay, the protein binds only to sequences that contain a nucleotide mismatch in a heteroduplex between mutant and wild-type sequences.

Mismatches, according to the present invention, are hybridized nucleic acid duplexes in which the two strands are not 100% complementary. Lack of total homology may be due to deletions insertions, inversions or substitutions. Mismatch detection can be used to detect point mutations in the gene or in its mRNA product. While these techniques are less sensitive than sequencing, they are simpler to perform on a large number of tumor samples. An example of a mismatch cleavage technique is the RNase protection method. In the practice of the present invention, the method involves the use of a labeled riboprobe which is complementary to the human wild-type BRCA2 gene coding sequence. The riboprobe and either mRNA or DNA isolated from the tumor tissue are annealed (hybridized) together and subsequently digested with the enzyme RNase A which is able to detect some mismatches in a duplex RNA structure. If a mismatch is detected by RNase A, it cleaves at the site of the mismatch. Thus, when the annealed RNA preparation is separated on an electrophoretic gel matrix, if a mismatch has been detected and cleaved by RNase A, an RNA product will be seen which is smaller than the full length duplex RNA for the riboprobe and the mRNA or DNA. The riboprobe need not be the full length of the BRCA2 mRNA or gene but can be a segment of either. If the riboprobe comprises only a segment of the BRCA2 mRNA or gene it will be desirable to use a number of these probes to screen the whole mRNA sequence for mismatches.

In similar fashion, DNA probes can be used to detect mismatches, through enzymatic or chemical cleavage. See,

e.g., Cotton et al., 1988; Shenk et al., 1975; Novack et al., 1986. Alternatively, mismatches can be detected by shifts in the electrophoretic mobility of mismatched duplexes relative to matched duplexes. See, e.g., Cariello, 1988. With either riboprobes or DNA probes, the cellular mRNA or DNA which might contain a mutation can be amplified using PCR (see below) before hybridization. Changes in DNA of the BRCA2 gene can also be detected using Southern hybridization, especially if the changes are gross rearrangements, such as deletions and insertions.

DNA sequences of the BRCA2 gene which have been amplified by use of PCR may also be screened using allele-specific probes. These probes are nucleic acid oligomers, each of which contains a region of the BRCA2 gene sequence harboring a known mutation. For example, one oligomer may be about 30 nucleotides in length, corresponding to a portion of the BRCA2 gene sequence. By use of a battery of such allele-specific probes, PCR amplification products can be screened to identify the presence of a previously identified mutation in the BRCA2 gene. Hybridization of allele-specific probes with amplified BRCA2 sequences can be performed, for example, on a nylon filter. Hybridization to a particular probe under stringent hybridization conditions indicates the presence of the same mutation in the tumor tissue as in the allele-specific probe.

The most definitive test for mutations in a candidate locus is to directly compare genomic BRCA2 sequences from cancer patients with those from a control population. Alternatively, one could sequence messenger RNA after amplification, e.g., by PCR, thereby eliminating the necessity of determining the exon structure of the candidate gene.

Mutations from cancer patients falling outside the coding region of BRCA2 can be detected by examining the non-coding regions, such as introns and regulatory sequences near or within the BRCA2 gene. An early indication that mutations in noncoding regions are important may come from Northern blot experiments that reveal messenger RNA molecules of abnormal size or abundance in cancer patients as compared to control individuals.

Alteration of BRCA2 mRNA expression can be detected by any techniques known in the art.

These include Northern blot analysis, PCR amplification and RNase protection. Diminished mRNA expression indicates an alteration of the wild-type BRCA2 gene. Alteration of wild-type BRCA2 genes can also be detected by screening for alteration of wild-type BRCA2 protein. For example, monoclonal antibodies immunoreactive with BRCA2 can be used to screen a tissue. Lack of cognate antigen would indicate a BRCA2 mutation. Antibodies specific for products of mutant alleles could also be used to detect mutant BRCA2 gene product. Such immunological assays can be done in any convenient formats known in the art. These include Western blots, immunohistochemical assays and ELISA assays. Any means for detecting an altered BRCA2 protein can be used to detect alteration of wild-type BRCA2 genes. Functional assays, such as protein binding determinations, can be used. In addition, assays can be used which detect BRCA2 biochemical function. Finding a mutant BRCA2 gene product indicates alteration of a wild-type BRCA2 gene.

Mutant BRCA2 genes or gene products can also be detected in other human body samples, such as serum, stool, urine and sputum. The same techniques discussed above for detection of mutant BRCA2 genes or gene products in tissues can be applied to other body samples. Cancer cells are sloughed off from tumors and appear in such body

samples. In addition, the BRCA2 gene product itself may be secreted into the extracellular space and found in these body samples even in the absence of cancer cells. By screening such body samples, a simple early diagnosis can be achieved for many types of cancers. In addition, the progress of chemotherapy or radiotherapy can be monitored more easily by testing such body samples for mutant BRCA2 genes or gene products.

The methods of diagnosis of the present invention are applicable to any tumor in which BRCA2 has a role in tumorigenesis. The diagnostic method of the present invention is useful for clinicians, so they can decide upon an appropriate course of treatment.

The primer pairs of the present invention are useful for determination of the nucleotide sequence of a particular BRCA2 allele using PCR. The pairs of single-stranded DNA primers can be annealed to sequences within or surrounding the BRCA2 gene on chromosome 13 in order to prime amplifying DNA synthesis of the BRCA2 gene itself. A complete set of these primers allows synthesis of all of the nucleotides of the BRCA2 gene coding sequences, i.e., the exons. The set of primers preferably allows synthesis of both intron and exon sequences. Allele-specific primers can also be used. Such primers anneal only to particular BRCA2 mutant alleles, and thus will only amplify a product in the presence of the mutant allele as a template.

In order to facilitate subsequent cloning of amplified sequences, primers may have restriction enzyme site sequences appended to their 5' ends. Thus, all nucleotides of the primers are derived from BRCA2 sequences or sequences adjacent to BRCA2, except for the few nucleotides necessary to form a restriction enzyme site. Such enzymes and sites are well known in the art. The primers themselves can be synthesized using techniques which are well known in the art. Generally, the primers can be made using oligonucleotide synthesizing machines which are commercially available. Given the sequence of the BRCA2 open reading frame shown in SEQ ID NO:1 and in FIG. 3, design of particular primers, in addition to those disclosed below, is well within the skill of the art.

The nucleic acid probes provided by the present invention are useful for a number of purposes. They can be used in Southern hybridization to genomic DNA and in the RNase protection method for detecting point mutations already discussed above. The probes can be used to detect PCR amplification products. They may also be used to detect mismatches with the BRCA2 gene or mRNA using other techniques.

It has been discovered that individuals with the wild-type BRCA2 gene do not have cancer which results from the BRCA2 allele. However, mutations which interfere with the function of the BRCA2 protein are involved in the pathogenesis of cancer. Thus, the presence of an altered (or a mutant) BRCA2 gene which produces a protein having a loss of function, or altered function, directly correlates to an increased risk of cancer. In order to detect a BRCA2 gene mutation, a biological sample is prepared and analyzed for a difference between the sequence of the BRCA2 allele being analyzed and the sequence of the wild-type BRCA2 allele. Mutant BRCA2 alleles can be initially identified by any of the techniques described above. The mutant alleles are then sequenced to identify the specific mutation of the particular mutant allele. Alternatively, mutant BRCA2 alleles can be initially identified by identifying mutant (altered) BRCA2 proteins, using conventional techniques. The mutant alleles are then sequenced to identify the specific mutation for each allele. The mutations, especially those

which lead to an altered function of the BRCA2 protein, are then used for the diagnostic and prognostic methods of the present invention.

Definitions

The present invention employs the following definitions: "Amplification of Polynucleotides" utilizes methods such as the polymerase chain reaction (PCR), ligation amplification (or ligase chain reaction, LCR) and amplification methods based on the use of Q-beta replicase. These methods are well known and widely practiced in the art. See, e.g. U.S. Pat. Nos. 4,683,195 and 4,683,202 and Innis et al., 1990 (for PCR); and Wu et al., 1989a (for LCR). Reagents and hardware for conducting PCR are commercially available. Primers useful to amplify sequences from the BRCA2 region are preferably complementary to, and hybridize specifically to sequences in the BRCA2 region or in regions that flank a target region therein. BRCA2 sequences generated by amplification may be sequenced directly. Alternatively, but less desirably, the amplified sequence(s) may be cloned prior to sequence analysis. A method for the direct cloning and sequence analysis of enzymatically amplified genomic segments has been described by Scharf, 1986.

"Analyte polynucleotide" and "analyte strand" refer to a single- or double-stranded polynucleotide which is suspected of containing a target sequence, and which may be present in a variety of types of samples, including biological samples.

"Antibodies." The present invention also provides polyclonal and/or monoclonal antibodies and fragments thereof, and immunologic binding equivalents thereof, which are capable of specifically binding to the BRCA2 polypeptides and fragments thereof or to polynucleotide sequences from the BRCA2 region, particularly from the BRCA2 locus or a portion thereof. The term "antibody" is used both to refer to a homogeneous molecular entity, or a mixture such as a serum product made up of a plurality of different molecular entities. Polypeptides may be prepared synthetically in a peptide synthesizer and coupled to a carrier molecule (e.g., keyhole limpet hemocyanin) and injected over several months into rabbits. Rabbit sera is tested for immunoreactivity to the BRCA2 polypeptide or fragment. Monoclonal antibodies may be made by injecting mice with the protein polypeptides, fusion proteins or fragments thereof. Monoclonal antibodies will be screened by ELISA and tested for specific immunoreactivity with BRCA2 polypeptide or fragments thereof. See, Harlow & Lane, 1988. These antibodies will be useful in assays as well as pharmaceuticals.

Once a sufficient quantity of desired polypeptide has been obtained, it may be used for various purposes. A typical use is the production of antibodies specific for binding. These antibodies may be either polyclonal or monoclonal, and may be produced by in vitro or in vivo techniques well known in the art. For production of polyclonal antibodies, an appropriate target immune system, typically mouse or rabbit, is selected. Substantially purified antigen is presented to the immune system in a fashion determined by methods appropriate for the animal and by other parameters well known to immunologists. Typical sites for injection are in footpads, intramuscularly, intraperitoneally, or intradermally. Of course, other species may be substituted for mouse or rabbit. Polyclonal antibodies are then purified using techniques known in the art, adjusted for the desired specificity.

An immunological response is usually assayed with an immunoassay. Normally, such immunoassays involve some purification of a source of antigen, for example, that produced by the same cells and in the same fashion as the antigen. A variety of immunoassay methods are well known in the art. See, e.g., Harlow & Lane, 1988, or Goding, 1986.

6,033,857

17

Monoclonal antibodies with affinities of 10^{-8} M⁻¹ or preferably 10^{-9} to 10^{-10} M⁻¹ or stronger will typically be made by standard procedures as described, e.g., in Harlow & Lane, 1988 or Goding, 1986. Briefly, appropriate animals will be selected and the desired immunization protocol followed. After the appropriate period of time, the spleens of such animals are excised and individual spleen cells fused, typically, to immortalized myeloma cells under appropriate selection conditions. Thereafter, the cells are clonally separated and the supernatants of each clone tested for their production of an appropriate antibody specific for the desired region of the antigen.

Other suitable techniques involve in vitro exposure of lymphocytes to the antigenic polypeptides, or alternatively, to selection of libraries of antibodies in phage or similar vectors. See Huse et al. 1989. The polypeptides and antibodies of the present invention may be used with or without modification. Frequently, polypeptides and antibodies will be labeled by joining, either covalently or non-covalently, a substance which provides for a detectable signal. A wide variety of labels and conjugation techniques are known and are reported extensively in both the scientific and patent literature. Suitable labels include radionuclides, enzymes, substrates, cofactors, inhibitors, fluorescent agents, chemiluminescent agents, magnetic particles and the like. Patents teaching the use of such labels include U.S. Pat. Nos. 3,817,837; 3,850,752; 3,939,350; 3,996,345; 4,277,437; 4,275,149 and 4,366,241. Also, recombinant immunoglobulins may be produced (see U.S. Pat. No. 4,816,567).

"Binding partner" refers to a molecule capable of binding a ligand molecule with high specificity, as for example, an antigen and an antigen-specific antibody or an enzyme and its inhibitor. In general, the specific binding partners must bind with sufficient affinity to immobilize the analyte copy/complementary strand duplex (in the case of polynucleotide hybridization) under the isolation conditions. Specific binding partners are known in the art and include, for example, biotin and avidin or streptavidin, IgG and protein A, the numerous, known receptor-ligand couples, and complementary polynucleotide strands. In the case of complementary polynucleotide binding partners, the partners are normally at least about 15 bases in length, and may be at least 40 bases in length. The polynucleotides may be composed of DNA, RNA, or synthetic nucleotide analogs.

A "biological sample" refers to a sample of tissue or fluid suspected of containing an analyte polynucleotide or polypeptide from an individual including, but not limited to, e.g., plasma, serum, spinal fluid, lymph fluid, the external sections of the skin, respiratory, intestinal, and genitourinary tracts, tears, saliva, blood cells, tumors, organs, tissue and samples of in vitro cell culture constituents.

As used herein, the terms "diagnosing" or "prognosing," as used in the context of neoplasia, are used to indicate 1) the classification of lesions as neoplasia, 2) the determination of the severity of the neoplasia, or 3) the monitoring of the disease progression, prior to, during and after treatment.

"Encode". A polynucleotide is said to "encode" a polypeptide if, in its native state or when manipulated by methods well known to those skilled in the art, it can be transcribed and/or translated to produce the mRNA for and/or the polypeptide or a fragment thereof. The anti-sense strand is the complement of such a nucleic acid, and the encoding sequence can be deduced therefrom.

"Isolated" or "substantially pure". An "isolated" or "substantially pure" nucleic acid (e.g., an RNA, DNA or a mixed polymer) is one which is substantially separated from other cellular components which naturally accompany a native

18

human sequence or protein, e.g., ribosomes, polymerases, many other human genome sequences and proteins. The term embraces a nucleic acid sequence or protein which has been removed from its naturally occurring environment, and includes recombinant or cloned DNA isolates and chemically synthesized analogs or analogs biologically synthesized by heterologous systems.

"BRCA2 Allele" refers to normal alleles of the BRCA2 locus as well as alleles carrying variations that predispose individuals to develop cancer of many sites including, for example, breast, ovarian and stomach cancer. Such predisposing alleles are also called "BRCA2 susceptibility alleles".

"BRCA2 Locus," "BRCA2 Gene," "BRCA2 Nucleic Acids" or "BRCA2 Polynucleotide" each refer to polynucleotides, all of which are in the BRCA2 region, that are likely to be expressed in normal tissue, certain alleles of which predispose an individual to develop breast, ovarian and stomach cancers. Mutations at the BRCA2 locus may be involved in the initiation and/or progression of other types of tumors. The locus is indicated in part by mutations that predispose individuals to develop cancer. These mutations fall within the BRCA2 region described infra. The BRCA2 locus is intended to include coding sequences, intervening sequences and regulatory elements controlling transcription and/or translation. The BRCA2 locus is intended to include all allelic variations of the DNA sequence.

These terms, when applied to a nucleic acid, refer to a nucleic acid which encodes a BRCA2 polypeptide, fragment, homolog or variant, including, e.g., protein fusions or deletions. The nucleic acids of the present invention will possess a sequence which is either derived from, or substantially similar to a natural BRCA2-encoding gene or one having substantial homology with a natural BRCA2-encoding gene or a portion thereof. The coding sequence for a BRCA2 polypeptide is shown in SEQ ID NO:1 and FIG. 3, with the amino acid sequence shown in SEQ ID NO:2.

The polynucleotide compositions of this invention include RNA, cDNA, genomic DNA, synthetic forms, and mixed polymers, both sense and antisense strands, and may be chemically or biochemically modified or may contain non-natural or derivatized nucleotide bases, as will be readily appreciated by those skilled in the art. Such modifications include, for example, labels, methylation, substitution of one or more of the naturally occurring nucleotides with an analog, internucleotide modifications such as uncharged linkages (e.g., methyl phosphonates, phosphotriesters, phosphoamidates, carbamates, etc.), charged linkages (e.g., phosphorothioates, phosphorodithioates, etc.), pendent moieties (e.g., polypeptides), intercalators (e.g., acridine, psoralen, etc.), chelators, alkylators, and modified linkages (e.g., alpha anomeric nucleic acids, etc.). Also included are synthetic molecules that mimic polynucleotides in their ability to bind to a designated sequence via hydrogen bonding and other chemical interactions. Such molecules are known in the art and include, for example, those in which peptide linkages substitute for phosphate linkages in the backbone of the molecule.

The present invention provides recombinant nucleic acids comprising all or part of the BRCA2 region. The recombinant construct may be capable of replicating autonomously in a host cell. Alternatively, the recombinant construct may become integrated into the chromosomal DNA of the host cell. Such a recombinant polynucleotide comprises a polynucleotide of genomic, cDNA, semi-synthetic, or synthetic origin which, by virtue of its origin or manipulation, 1) is not

associated with all or a portion of a polynucleotide with which it is associated in nature; 2) is linked to a polynucleotide other than that to which it is linked in nature; or 3) does not occur in nature.

Therefore, recombinant nucleic acids comprising sequences otherwise not naturally occurring are provided by this invention. Although the wild-type sequence may be employed, it will often be altered, e.g., by deletion, substitution or insertion.

cDNA or genomic libraries of various types may be screened as natural sources of the nucleic acids of the present invention, or such nucleic acids may be provided by amplification of sequences resident in genomic DNA or other natural sources, e.g., by PCR. The choice of cDNA libraries normally corresponds to a tissue source which is abundant in MRNA for the desired proteins. Phage libraries are normally preferred, but other types of libraries may be used. Clones of a library are spread onto plates, transferred to a substrate for screening, denatured and probed for the presence of desired sequences.

The DNA sequences used in this invention will usually comprise at least about five codons (15 nucleotides), more usually at least about 7–15 codons, and most preferably, at least about 35 codons. One or more introns may also be present. This number of nucleotides is usually about the minimal length required for a successful probe that would hybridize specifically with a BRCA2-encoding sequence.

Techniques for nucleic acid manipulation are described generally, for example, in Sambrook et al., 1989 or Ausubel et al., 1992. Reagents useful in applying such techniques, such as restriction enzymes and the like, are widely known in the art and commercially available from such vendors as New England BioLabs, Boehringer Mannheim, Amersham, Promega Biotec, U. S. Biochemicals, New England Nuclear, and a number of other sources. The recombinant nucleic acid sequences used to produce fusion proteins of the present invention may be derived from natural or synthetic sequences. Many natural gene sequences are obtainable from various cDNA or from genomic libraries using, appropriate probes. See, GenBank, National Institutes of Health.

“BRCA2 Region” refers to a portion of human chromosome 13 bounded by the markers tdj3820 and YS-G-B10T. This region contains the BRCA2 locus, including the BRCA2 gene.

As used herein, the terms “BRCA2 locus,” “BRCA2 allele” and “BRCA2 region” all refer to the double-stranded DNA comprising the locus, allele, or region, as well as either of the single-stranded DNAs comprising the locus, allele or region.

As used herein, a “portion” of the BRCA2 locus or region or allele is defined as having a minimal size of at least about eight nucleotides, or preferably about 15 nucleotides, or more preferably at least about 25 nucleotides, and may have a minimal size of at least about 40 nucleotides.

“BRCA2 protein” or “BRCA2 polypeptide” refer to a protein or polypeptide encoded by the BRCA2 locus, variants or fragments thereof. The term “polypeptide” refers to a polymer of amino acids and its equivalent and does not refer to a specific length of the product; thus, peptides, oligopeptides and proteins are included within the definition of a polypeptide. This term also does not refer to, or exclude modifications of the polypeptide, for example, glycosylations, acetylations, phosphorylations, and the like. Included within the definition are, for example, polypeptides containing one or more analogs of an amino acid (including, for example, unnatural amino acids, etc.), polypeptides with substituted linkages as well as other modifications known in

the art, both naturally and non-naturally occurring. Ordinarily, such polypeptides will be at least about 50% homologous to the native BRCA2 sequence, preferably in excess of about 90%, and more preferably at least about 95% homologous. Also included are proteins encoded by DNA which hybridize under high or low stringency conditions, to BRCA2-encoding nucleic acids and closely related polypeptides or proteins retrieved by antisera to the BRCA2 protein (s).

The length of polypeptide sequences compared for homology will generally be at least about 16 amino acids, usually at least about 20 residues, more usually at least about 24 residues, typically at least about 28 residues, and preferably more than about 35 residues. “Operably linked” refers to a juxtaposition wherein the components so described are in a relationship permitting them to function in their intended manner. For instance, a promoter is operably linked to a coding sequence if the promoter affects its transcription or expression.

“Probes”. Polynucleotide polymorphisms associated with BRCA2 alleles which predispose to certain cancers or are associated with most cancers are detected by hybridization with a polynucleotide probe which forms a stable hybrid with that of the target sequence, under stringent to moderately stringent hybridization and wash conditions. If it is expected that the probes will be perfectly complementary to the target sequence, stringent conditions will be used. Hybridization stringency may be lessened if some mismatching is expected, for example, if variants are expected with the result that the probe will not be completely complementary. Conditions are chosen which rule out nonspecific/adventitious bindings, that is, which minimize noise. Since such indications identify neutral DNA polymorphisms as well as mutations, these indications need further analysis to demonstrate detection of a BRCA2 susceptibility allele.

Probes for BRCA2 alleles may be derived from the sequences of the BRCA2 region or its cDNAs. The probes may be of any suitable length, which span all or a portion of the BRCA2 region, and which allow specific hybridization to the BRCA2 region. If the target sequence contains a sequence identical to that of the probe, the probes may be short, e.g., in the range of about 8–30 base pairs, since the hybrid will be relatively stable under even stringent conditions. If some degree of mismatch is expected with the probe, i.e., if it is suspected that the probe will hybridize to a variant region, a longer probe may be employed which hybridizes to the target sequence with the requisite specificity.

The probes will include an isolated polynucleotide attached to a label or reporter molecule and may be used to isolate other polynucleotide sequences, having sequence similarity by standard methods. For techniques for preparing and labeling probes see, e.g., Sambrook et al., 1989 or Ausubel et al., 1992. Other similar polynucleotides may be selected by using homologous polynucleotides. Alternatively, polynucleotides encoding these or similar polypeptides may be synthesized or selected by use of the redundancy in the genetic code. Various codon substitutions may be introduced, e.g., by silent changes (thereby producing various restriction sites) or to optimize expression for a particular system. Mutations may be introduced to modify the properties of the polypeptide, perhaps to change ligand-binding affinities, interchain affinities, or the polypeptide degradation or turnover rate.

Probes comprising synthetic oligonucleotides or other polynucleotides of the present invention may be derived from naturally occurring or recombinant single- or double-

6,033,857

21

stranded polynucleotides, or be chemically synthesized. Probes may also be labeled by nick translation, Klenow fill-in reaction, or other methods known in the art.

Portions of the polynucleotide sequence having at least about eight nucleotides, usually at least about 15 nucleotides, and fewer than about 6 kb, usually fewer than about 1.0 kb, from a polynucleotide sequence encoding BRCA2 are preferred as probes. The probes may also be used to determine whether mRNA encoding BRCA2 is present in a cell or tissue.

"Protein modifications or fragments" are provided by the present invention for BRCA2 polypeptides or fragments thereof which are substantially homologous to primary structural sequence but which include, e.g., in vivo or in vitro chemical and biochemical modifications or which incorporate unusual amino acids. Such modifications include, for example, acetylation, carboxylation, phosphorylation, glycosylation, ubiquitination, labeling, e.g., with radionuclides, and various enzymatic modifications, as will be readily appreciated by those well skilled in the art. A variety of methods for labeling polypeptides and of substituents or labels useful for such purposes are well known in the art, and include radioactive isotopes such as ^{32}P , ligands which bind to labeled antigens (e.g., antibodies), fluorophores, chemiluminescent agents, enzymes, and antigens which can serve as specific binding pair members for a labeled ligand. The choice of label depends on the sensitivity required, ease of conjugation with the primer, stability requirements, and available instrumentation. Methods of labeling polypeptides are well known in the art. See, e.g., Sambrook et al., 1989 or Ausubel et al., 1992.

Besides substantially full-length polypeptides, the present invention provides for biologically active fragments of the polypeptides. Significant biological activities include ligand-binding, immunological activity and other biological activities characteristic of BRCA2 polypeptides. Immunological activities include both immunogenic function in a target immune system, as well as sharing of immunological epitopes for binding, serving as either a competitor or substitute antigen for an epitope of the BRCA2 protein. As used herein, "epitope" refers to an antigenic determinant of a polypeptide. An epitope could comprise three amino acids in a spatial conformation which is unique to the epitope. Generally, an epitope consists of at least five such amino acids, and more usually consists of at least 8–10 such amino acids. Methods of determining the spatial conformation of such amino acids are known in the art.

For immunological purposes, tandem-repeat polypeptide segments may be used as immunogens, thereby producing highly antigenic proteins. Alternatively, such polypeptides will serve as highly efficient competitors for specific binding. Production of antibodies specific for BRCA2 polypeptides or fragments thereof is described below.

The present invention also provides for fusion polypeptides, comprising BRCA2 polypeptides and fragments. Homologous polypeptides may be fusions between two or more BRCA2 polypeptide sequences or between the sequences of BRCA2 and a related protein. Likewise, heterologous fusions may be constructed which would exhibit a combination of properties or activities of the derivative proteins. For example, ligand-binding or other domains may be "swapped" between different new fusion polypeptides or fragments. Such homologous or heterologous fusion polypeptides may display, for example, altered strength or specificity of binding. Fusion partners include immunoglobulins, bacterial β -galactosidase, trpE, protein A,

22

β -lactamase, alpha amylase, alcohol dehydrogenase and yeast alpha mating factor. See, e.g., Godowski et al., 1988.

Fusion proteins will typically be made by either recombinant nucleic acid methods, as described below, or may be chemically synthesized. Techniques for the synthesis of polypeptides are described, for example, in Merrifield, 1963.

"Protein purification" refers to various methods for the isolation of the BRCA2 polypeptides from other biological material, such as from cells transformed with recombinant nucleic acids encoding BRCA2, and are well known in the art. For example, such polypeptides may be purified by immunoaffinity chromatography employing, e.g., the antibodies provided by the present invention. Various methods of protein purification are well known in the art, and include those described in Deutscher, 1990 and Scopes, 1982.

The terms "isolated", "substantially pure", and "substantially homogeneous" are used interchangeably to describe a protein or polypeptide which has been separated from components which accompany it in its natural state. A monomeric protein is substantially pure when at least about 60 to 75% of a sample exhibits a single polypeptide sequence. A substantially pure protein will typically comprise about 60 to 90% w/w of a protein sample, more usually about 95%, and preferably will be over about 99% pure. Protein purity or homogeneity may be indicated by a number of means well known in the art, such as polyacrylamide gel electrophoresis of a protein sample, followed by visualizing a single polypeptide band upon staining the gel. For certain purposes, higher resolution may be provided by using HPLC or other means well known in the art which are utilized for purification.

A BRCA2 protein is substantially free of naturally associated components when it is separated from the native contaminants which accompany it in its natural state. Thus, a polypeptide which is chemically synthesized or synthesized in a cellular system different from the cell from which it naturally originates will be substantially free from its naturally associated components. A protein may also be rendered substantially free of naturally associated components by isolation, using protein purification techniques well known in the art.

A polypeptide produced as an expression product of an isolated and manipulated genetic sequence is an "isolated polypeptide," as used herein, even if expressed in a homologous cell type.

Synthetically made forms or molecules expressed by heterologous cells are inherently isolated molecules. "Recombinant nucleic acid" is a nucleic acid which is not naturally occurring, or which is made by the artificial combination of two otherwise separated segments of sequence. This artificial combination is often accomplished by either chemical synthesis means, or by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques. Such is usually done to replace a codon with a redundant codon encoding the same or a conservative amino acid, while typically introducing or removing a sequence recognition site. Alternatively, it is performed to join together nucleic acid segments of desired functions to generate a desired combination of functions.

"Regulatory sequences" refers to those sequences normally within 100 kb of the coding region of a locus, but they may also be more distant from the coding region, which affect the expression of the gene (including transcription of the gene, and translation, splicing, stability or the like of the messenger RNA).

"Substantial homology or similarity". A nucleic acid or fragment thereof is "substantially homologous" ("or sub-

23

stantially similar”) to another if, when optimally aligned (with appropriate nucleotide insertions or deletions) with the other nucleic acid (or its complementary strand), there is nucleotide sequence identity in at least about 60% of the nucleotide bases, usually at least about 70%, more usually at least about 80%, preferably at least about 90%, and more preferably at least about 95–98% of the nucleotide bases.

Alternatively, substantial homology or (similarity) exists when a nucleic acid or fragment thereof will hybridize to another nucleic acid (or a complementary strand thereof) under selective hybridization conditions, to a strand, or to its complement. Selectivity of hybridization exists when hybridization which is substantially more selective than total lack of specificity occurs. Typically, selective hybridization will occur when there is at least about 55% homology over a stretch of at least about 14 nucleotides, preferably at least about 65%, more preferably at least about 75%, and most preferably at least about 90%. See, Kanehisa, 1984. The length of homology comparison, as described, may be over longer stretches, and in certain embodiments will often be over a stretch of at least about nine nucleotides, usually at least about 20 nucleotides, more usually at least about 24 nucleotides, typically at least about 28 nucleotides, more typically at least about 32 nucleotides, and preferably at least about 36 or more nucleotides.

Nucleic acid hybridization will be affected by such conditions as salt concentration, temperature, or organic solvents, in addition to the base composition, length of the complementary strands, and the number of nucleotide base mismatches between the hybridizing nucleic acids, as will be readily appreciated by those skilled in the art. Stringent temperature conditions will generally include temperatures in excess of 30° C., typically in excess of 37° C., and preferably in excess of 45° C. Stringent salt conditions will ordinarily be less than 1000 mM, typically less than 500 mM, and preferably less than 200 mM. However, the combination of parameters is much more important than the measure of any single parameter. See, e.g., Wetmur & Davidson, 1968.

Probe sequences may also hybridize specifically to duplex DNA under certain conditions to form triplex or other higher order DNA complexes. The preparation of such probes and suitable hybridization conditions are well known in the art.

The terms “substantial homology” or “substantial identity”, when referring to polypeptides, indicate that the polypeptide or protein in question exhibits at least about 30% identity with an entire naturally-occurring protein or a portion thereof, usually at least about 70% identity, and preferably at least about 95% identity.

“Substantially similar function” refers to the function of a modified nucleic acid or a modified protein, with reference to the wild-type BRCA2 nucleic acid or wild-type BRCA2 polypeptide. The modified polypeptide will be substantially homologous to the wild-type BRCA2 polypeptide and will have substantially the same function. The modified polypeptide may have an altered amino acid sequence and/or may contain modified amino acids. In addition to the similarity of function, the modified polypeptide may have other useful properties, such as a longer half-life. The similarity of function (activity) of the modified polypeptide may be substantially the same as the activity of the wild-type BRCA2 polypeptide. Alternatively, the similarity of function (activity) of the modified polypeptide may be higher than the activity of the wild-type BRCA2 polypeptide. The modified polypeptide is synthesized using conventional techniques, or is encoded by a modified nucleic acid and produced using conventional techniques. The modified

24

nucleic acid is prepared by conventional techniques. A nucleic acid with a function substantially similar to the wild-type BRCA2 gene function produces the modified protein described above.

Homology, for polypeptides, is typically measured using sequence analysis software. See, e.g., the Sequence Analysis Software Package of the Genetics Computer Group, University of Wisconsin Biotechnology Center, 910 University Avenue, Madison, Wis. 53705. Protein analysis software matches similar sequences using measure of homology assigned to various substitutions, deletions and other modifications. Conservative substitutions typically include substitutions within the following groups: glycine, alanine; valine, isoleucine, leucine; aspartic acid, glutamic acid; asparagine, glutamine; serine, threonine; lysine, arginine; and phenylalanine, tyrosine.

A polypeptide "fragment," "portion" or "segment" is a stretch of amino acid residues of at least about five to seven contiguous amino acids, often at least about seven to nine contiguous amino acids, typically at least about nine to 13 contiguous amino acids and, most preferably, at least about 20 to 30 or more contiguous amino acids. The polypeptides of the present invention, if soluble, may be coupled to a solid-phase support, e.g., nitrocellulose, nylon, column packing materials (e.g., Sepharose beads), magnetic beads, glass wool, plastic, metal, polymer gels, cells, or other substrates. Such supports may take the form, for example, of beads, wells, dipsticks, or membranes.

“Target region” refers to a region of the nucleic acid which is amplified and/or detected. The term “target 30 sequence” refers to a sequence with which a probe or primer will form a stable hybrid under desired conditions.

The practice of the present invention employs, unless otherwise indicated, conventional techniques of chemistry, molecular biology, microbiology, recombinant DNA, genetics, and immunology. See, e.g., Maniatis et al., 1982; Sambrook et al., 1989; Ausubel et al., 1992; Glover, 1985; Anand, 1992; Guthrie & Fink, 1991. A general discussion of techniques and materials for human gene mapping, including mapping of human chromosome 13, is provided, e.g., in White and Lalouel, 1988.

Preparation of recombinant or chemically synthesized nucleic acids; vectors, transformation, host cells

Large amounts of the polynucleotides of the present invention may be produced by replication in a suitable host cell. Natural or synthetic polynucleotide fragments coding for a desired fragment will be incorporated into recombinant polynucleotide constructs, usually DNA constructs, capable of introduction into and replication in a prokaryotic or eukaryotic cell. Usually the polynucleotide constructs will be suitable for replication in a unicellular host, such as yeast or bacteria, but may also be intended for introduction to (with and without integration within the genome) cultured mammalian or plant or other eukaryotic cell lines. The purification of nucleic acids produced by the methods of the present invention is described, e.g., in Sambrook et al., 1989 or Ausubel et al., 1992.

The polynucleotides of the present invention may also be produced by chemical synthesis, e.g., by the phosphoramidite method described by Beaucage & Caruthers, 1981 or the triester method according to Matteucci and Caruthers, 1981, and may be performed on commercial, automated oligonucleotide synthesizers. A double-stranded fragment may be obtained from the single-stranded product of chemical synthesis either by synthesizing the complementary strand and annealing the strands together under appropriate conditions or by adding the complementary strand using DNA polymerase with an appropriate primer sequence.

Polynucleotide constructs prepared for introduction into a prokaryotic or eukaryotic host may comprise a replication system recognized by the host, including the intended polynucleotide fragment encoding the desired polypeptide, and will preferably also include transcription and translational initiation regulatory sequences operably linked to the polypeptide encoding segment. Expression vectors may include, for example, an origin of replication or autonomously replicating sequence (ARS) and expression control sequences, a promoter, an enhancer and necessary processing information sites, such as ribosome-binding sites, RNA splice sites, polyadenylation sites, transcriptional terminator sequences, and mRNA stabilizing sequences. Secretion signals may also be included where appropriate, whether from a native BRCA2 protein or from other receptors or from secreted polypeptides of the same or related species, which allow the protein to cross and/or lodge in cell membranes, and thus attain its functional topology, or be secreted from the cell. Such vectors may be prepared by means of standard recombinant techniques well known in the art and discussed, for example, in Sambrook et al, 1989 or Ausubel et al. 1992.

An appropriate promoter and other necessary vector sequences will be selected so as to be functional in the host, and may include, when appropriate, those naturally associated with BRCA2 genes. Examples of workable combinations of cell lines and expression vectors are described in Sambrook et al., 1989 or Ausubel et al., 1992; see also, e.g., Metzger et al., 1988. Many useful vectors are known in the art and may be obtained from such vendors as Stratagene, New England BioLabs, Promega Biotech, and others. Promoters such as the trp, lac and phage promoters, tRNA promoters and glycolytic enzyme promoters may be used in prokaryotic hosts. Useful yeast promoters include promoter regions for metallothionein, 3-phosphoglycerate kinase or other glycolytic enzymes such as enolase or glyceraldehyde-3-phosphate dehydrogenase, enzymes responsible for maltose and galactose utilization, and others. Vectors and promoters suitable for use in yeast expression are further described in Hitzeman et al., EP 73,675A. Appropriate non-native mammalian promoters might include the early and late promoters from SV40 (Fiers et al, 1978) or promoters derived from murine Moloney leukemia virus, mouse tumor virus, avian sarcoma viruses, adenovirus II, bovine papilloma virus or polyoma. In addition, the construct may be joined to an amplifiable gene (e.g., DHFR) so that multiple copies of the gene may be made. For appropriate enhancer and other expression control sequences, see also *Enhancers and Eukaryotic Gene Expression*, Cold Spring Harbor Press, Cold Spring Harbor, N.Y. (1983).

While such expression vectors may replicate autonomously, they may also replicate by being inserted into the genome of the host cell, by methods well known in the art.

Expression and cloning vectors will likely contain a selectable marker, a gene encoding a protein necessary for survival or growth of a host cell transformed with the vector. The presence of this gene ensures growth of only those host cells which express the inserts. Typical selection genes encode proteins that a) confer resistance to antibiotics or other toxic substances, e.g. ampicillin, neomycin, methotrexate, etc.; b) complement auxotrophic deficiencies, or c) supply critical nutrients not available from complex media, e.g., the gene encoding D-alanine racemase for *Bacilli*. The choice of the proper selectable marker will depend on the host cell, and appropriate markers for different hosts are well known in the art.

The vectors containing the nucleic acids of interest can be transcribed in vitro, and the resulting RNA introduced into

the host cell by well-known methods, e.g., by injection (see, Kubo et al. 1988), or the vectors can be introduced directly into host cells by methods well known in the art, which vary depending on the type of cellular host, including electroporation; transfection employing calcium chloride, rubidium chloride, calcium phosphate, DEAE-dextran, or other substances; microprojectile bombardment; lipofection; infection (where the vector is an infectious agent, such as a retroviral genome); and other methods. See generally, Sambrook et al, 1989 and Ausubel et al, 1992. The introduction of the polynucleotides into the host cell by any method known in the art, including inter alia, those described above, will be referred to herein as "transformation." The cells into which have been introduced nucleic acids described above are meant to also include the progeny of such cells.

Large quantities of the nucleic acids and polypeptides of the present invention may be prepared by expressing the BRCA2 nucleic acids or portions thereof in vectors or other expression vehicles in compatible prokaryotic or eukaryotic host cells. The most commonly used prokaryotic hosts are strains of *Escherichia coli*, although other prokaryotes, such as *Bacillus subtilis* or *Pseudomonas* may also be used.

Mammalian or other eukaryotic host cells, such as those of yeast, filamentous fungi, plant, insect, or amphibian or avian species, may also be useful for production of the proteins of the present invention. Propagation of mammalian cells in culture is per se well known. See, Jakoby and Pastan, 1979. Examples of commonly used mammalian host cell lines are VERO and HeLa cells, Chinese hamster ovary (CHO) cells, and WI38, BHK, and COS cell lines, although it will be appreciated by the skilled practitioner that other cell lines may be appropriate, e.g., to provide higher expression, desirable glycosylation patterns, or other features.

Clones are selected by using markers depending on the mode of the vector construction. The marker may be on the same or a different DNA molecule, preferably the same DNA molecule. In prokaryotic hosts, the transformant may be selected, e.g., by resistance to ampicillin, tetracycline or other antibiotics. Production of a particular product based on temperature sensitivity may also serve as an appropriate marker.

Prokaryotic or eukaryotic cells transformed with the polynucleotides of the present invention will be useful not only for the production of the nucleic acids and polypeptides of the present invention, but also, for example in studying the characteristics of BRCA2 polypeptides.

Antisense polynucleotide sequences are useful in preventing or diminishing the expression of the BRCA2 locus, as will be appreciated by those skilled in the art. For example, polynucleotide vectors containing all or a portion of the BRCA2 locus or other sequences from the BRCA2 region (particularly those flanking the BRCA2 locus) may be placed under the control of a promoter in an antisense orientation and introduced into a cell. Expression of such an antisense construct within a cell will interfere with BRCA2 transcription and/or translation and/or replication.

The probes and primers based on the BRCA2 gene sequences disclosed herein are used to identify homologous BRCA2 gene sequences and proteins in other species. These BRCA2 gene sequences and proteins are used in the diagnostic/prognostic, therapeutic and drug screening methods described herein for the species from which they have been isolated.

Methods of Use: Nucleic Acid Diagnosis and Diagnostic Kits

In order to detect the presence of a BRCA2 allele predisposing an individual to cancer, a biological sample such

6,033,857

27

as blood is prepared and analyzed for the presence or absence of susceptibility alleles of BRCA2. In order to detect the presence of neoplasia, the progression toward malignancy of a precursor lesion, or as a prognostic indicator, a biological sample of the lesion is prepared and analyzed for the presence or absence of mutant alleles of BRCA2. Results of these tests and interpretive information are returned to the health care provider for communication to the tested individual. Such diagnoses may be performed by diagnostic laboratories, or, alternatively, diagnostic kits are manufactured and sold to health care providers or to private individuals for self-diagnosis.

Initially, the screening method involves amplification of the relevant BRCA2 sequences. In another preferred embodiment of the invention, the screening method involves a non-PCR based strategy. Such screening methods include two-step label amplification methodologies that are well known in the art. Both PCR and non-PCR based screening strategies can detect target sequences with a high level of sensitivity.

The most popular method used today is target amplification. Here, the target nucleic acid sequence is amplified with polymerases. One particularly preferred method using polymerase-driven amplification is the polymerase chain reaction (PCR). The polymerase chain reaction and other polymerase-driven amplification assays can achieve over a million-fold increase in copy number through the use of polymerase-driven amplification cycles. Once amplified, the resulting nucleic acid can be sequenced or used as a substrate for DNA probes.

When the probes are used to detect the presence of the target sequences (for example, in screening for cancer susceptibility), the biological sample to be analyzed, such as blood or serum, may be treated, if desired, to extract the nucleic acids. The sample nucleic acid may be prepared in various ways to facilitate detection of the target sequence; e.g. denaturation, restriction digestion, electrophoresis or dot blotting. The targeted region of the analyte nucleic acid usually must be at least partially single-stranded to form hybrids with the targeting sequence of the probe. If the sequence is naturally single-stranded, denaturation will not be required. However, if the sequence is double-stranded, the sequence will probably need to be denatured. Denaturation can be carried out by various techniques known in the art.

Analyte nucleic acid and probe are incubated under conditions which promote stable hybrid formation of the target sequence in the probe with the putative targeted sequence in the analyte.

The region of the probes which is used to bind to the analyte can be made completely complementary to the targeted region of human chromosome 13. Therefore, high stringency conditions are desirable in order to prevent false positives. However, conditions of high stringency are used only if the probes are complementary to regions of the chromosome which are unique in the genome. The stringency of hybridization is determined by a number of factors during hybridization and during the washing procedure, including temperature, ionic strength, base composition, probe length, and concentration of formamide. These factors are outlined in, for example, Maniatis et al., 1982 and Sambrook et al., 1989. Under certain circumstances, the formation of higher order hybrids, such as triplexes, quadruplexes, etc., may be desired to provide the means of detecting target sequences.

Detection, if any, of the resulting hybrid is usually accomplished by the use of labeled probes. Alternatively, the probe

28

may be unlabeled, but may be detectable by specific binding with a ligand which is labeled, either directly or indirectly. Suitable labels, and methods for labeling probes and ligands are known in the art, and include, for example, radioactive labels which may be incorporated by known methods (e.g., nick translation, random priming or kinasin), biotin, fluorescent groups, chemiluminescent groups (e.g., dioxetanes, particularly triggered dioxetanes), enzymes, antibodies and the like. Variations of this basic scheme are known in the art, and include those variations that facilitate separation of the hybrids to be detected from extraneous materials and/or that amplify the signal from the labeled moiety. A number of these variations are reviewed in, e.g., Matthews & Kricka, 1988, Landegren et al., 1988; Mittlin, 1989; U.S. Pat. No. 4,868,105, and in EPO Publication No. 225,807.

As noted above, non-PCR based screening assays are also contemplated in this invention. An exemplary non-PCR based procedure is provided in Example 6. This procedure hybridizes a nucleic acid probe (or an analog such as a methyl phosphonate backbone replacing the normal phosphodiester), to the low level DNA target. This probe may have an enzyme covalently linked to the probe, such that the covalent linkage does not interfere with the specificity of the hybridization. This enzyme-probe-conjugate-target nucleic acid complex can then be isolated away from the free probe enzyme conjugate and a substrate is added for enzyme detection. Enzymatic activity is observed as a change in color development or luminescent output resulting in a 10^3 – 10^6 increase in sensitivity. For an example relating to preparation of oligodeoxynucleotide-alkaline phosphatase conjugates and their use as hybridization probes, see Jablonski et al., 1986.

Two-step label amplification methodologies are known in the art. These assays work on the principle that a small ligand (such as digoxigenin, biotin, or the like) is attached to a nucleic acid probe capable of specifically binding BRCA2. Exemplary probes can be developed on the basis of the sequence set forth in SEQ ID NO:1 and FIG. 3 of this patent application. Allele-specific probes are also contemplated within the scope of this example, and exemplary allele specific probes include probes encompassing the predisposing mutations described below, including those described in Table 2.

In one example, the small ligand attached to the nucleic acid probe is specifically recognized by an antibody-enzyme conjugate. In one embodiment of this example, digoxigenin is attached to the nucleic acid probe. Hybridization is detected by an antibody-alkaline phosphatase conjugate which turns over a chemiluminescent substrate. For methods for labeling nucleic acid probes according to this embodiment see Martin et al., 1990. In a second example, the small ligand is recognized by a second ligand-enzyme conjugate that is capable of specifically complexing to the first ligand. A well known embodiment of this example is the biotin-avidin type of interactions.

For methods for labeling nucleic acid probes and their use in biotin-avidin based assays see Rigby et al., 1977 and Nguven et al., 1992.

It is also contemplated within the scope of this invention that the nucleic acid probe assays of this invention will employ a cocktail of nucleic acid probes capable of detecting BRCA2. Thus, in one example to detect the presence of BRCA2 in a cell sample, more than one probe complementary to BRCA2 is employed and in particular the number of different probes is alternatively 2, 3, or 5 different nucleic acid probe sequences. In another example, to detect the presence of mutations in the BRCA2 gene sequence in a

6,033,857

29

patient, more than one probe complementary to BRCA2 is employed where the cocktail includes probes capable of binding to the allele-specific mutations identified in populations of patients with alterations in BRCA2. In this embodiment, any number of probes can be used, and will preferably include probes corresponding to the major gene mutations identified as predisposing an individual to breast cancer. Some candidate probes contemplated within the scope of the invention include probes that include the allele-specific mutations described below and those that have the BRCA2 regions shown in SEQ ID NO:1 and FIG. 3, both 5' and 3' to the mutation site.

Methods of Use: Peptide Diagnosis and Diagnostic Kits

The neoplastic condition of lesions can also be detected on the basis of the alteration of wild-type BRCA2 polypeptide. Such alterations can be determined by sequence analysis in accordance with conventional techniques. More preferably, antibodies (polyclonal or monoclonal) are used to detect differences in, or the absence of BRCA2 peptides. The antibodies may be prepared as discussed above under the heading "Antibodies" and as further shown in Examples 9 and 10. Other techniques for raising and purifying antibodies are well known in the art and any such techniques may be chosen to achieve the preparations claimed in this invention. In a preferred embodiment of the invention, antibodies will immunoprecipitate BRCA2 proteins from solution as well as react with BRCA2 protein on Western or immunoblots of polyacrylamide gels. In another preferred embodiment, antibodies will detect BRCA2 proteins in paraffin or frozen tissue sections, using immunocytochemical techniques.

Preferred embodiments relating to methods for detecting BRCA2 or its mutations include enzyme linked immunosorbent assays (ELISA), radioimmunoassays (RIA), immunoradiometric assays (IRMA) and immunoenzymatic assays (IEMA), including sandwich assays using monoclonal and/or polyclonal antibodies. Exemplary sandwich assays are described by David et al. in U.S. Pat. Nos. 4,376,110 and 4,486,530, hereby incorporated by reference, and exemplified in Example 9.

Methods of Use: Drug Screening

This invention is particularly useful for screening compounds by using the BRCA2 polypeptide or binding fragment thereof in any of a variety of drug screening techniques.

The BRCA2 polypeptide or fragment employed in such a test may either be free in solution, affixed to a solid support, or home on a cell surface. One method of drug screening utilizes eucaryotic or procaryotic host cells which are stably transformed with recombinant polynucleotides expressing the polypeptide or fragment, preferably in competitive binding assays. Such cells, either in viable or fixed form, can be used for standard binding assays. One may measure, for example, for the formation of complexes between a BRCA2 polypeptide or fragment and the agent being tested, or examine the degree to which the formation of a complex between a BRCA2 polypeptide or fragment and a known ligand is interfered with by the agent being tested.

Thus, the present invention provides methods of screening for drugs comprising contacting such an agent with a BRCA2 polypeptide or fragment thereof and assaying (i) for the presence of a complex between the agent and the BRCA2 polypeptide or fragment, or (ii) for the presence of a complex between the BRCA2 polypeptide or fragment and a ligand, by methods well known in the art. In such competitive binding assays the BRCA2 polypeptide or fragment is typically labeled. Free BRCA2 polypeptide or fragment is

30

separated from that present in a protein:protein complex, and the amount of free (i.e., uncomplexed) label is a measure of the binding of the agent being tested to BRCA2 or its interference with BRCA2:ligand binding, respectively.

Another technique for drug screening provides high throughput screening for compounds having suitable binding affinity to the BRCA2 polypeptides and is described in detail in Geysen, PCT published application WO 84/03564, published on Sep. 13, 1984. Briefly stated, large numbers of different small peptide test compounds are synthesized on a solid substrate, such as plastic pins or some other surface. The peptide test compounds are reacted with BRCA2 polypeptide and washed. Bound BRCA2 polypeptide is then detected by methods well known in the art. Purified BRCA2 can be coated directly onto plates for use in the aforementioned drug screening techniques. However non-neutralizing antibodies to the polypeptide can be used to capture antibodies to immobilize the BRCA2 polypeptide on the solid phase.

20 This invention also contemplates the use of competitive drug screening assays in which neutralizing antibodies capable of specifically binding the BRCA2 polypeptide compete with a test compound for binding to the BRCA2 polypeptide or fragments thereof. In this manner, the anti-
25 bodies can be used to detect the presence of any peptide which shares one or more antigenic determinants of the BRCA2 polypeptide.

A further technique for drug screening involves the use of host eukaryotic cell lines or cells (such as described above) which have a nonfunctional BRCA2 gene. These host cell lines or cells are defective at the BRCA2 polypeptide level. The host cell lines or cells are grown in the presence of drug compound. The rate of growth of the host cells is measured to determine if the compound is capable of regulating the growth of BRCA2 defective cells.

Methods of Use: Rational Drug Design

The goal of rational drug design is to produce structural analogs of biologically active polypeptides of interest or of small molecules with which they interact (e.g., agonists, antagonists, inhibitors) in order to fashion drugs which are, for example, more active or stable forms of the polypeptide, or which, e.g., enhance or interfere with the function of a polypeptide *in vivo*. See, e.g., Hodgson, 1991. In one approach, one first determines the three-dimensional structure of a protein of interest (e.g., BRCA2 polypeptide) or, for example, of the BRCA2-receptor or ligand complex, by x-ray crystallography, by computer modeling or most typically, by a combination of approaches. Less often, useful information regarding the structure of a polypeptide may be gained by modeling based on the structure of homologous proteins. An example of rational drug design is the development of HIV protease inhibitors (Erickson et al., 1990). In addition, peptides (e.g., BRCA2 polypeptide) are analyzed by an alanine scan (Wells, 1991). In this technique, an amino acid residue is replaced by Ala, and its effect on the peptide's activity is determined. Each of the amino acid residues of the peptide is analyzed in this manner to determine the important regions of the peptide.

It is also possible to isolate a target-specific antibody, 60 selected by a functional assay, and then to solve its crystal structure. In principle, this approach yields a pharmacore upon which subsequent drug design can be based. It is possible to bypass protein crystallography altogether by generating anti-idiotypic antibodies (anti-ids) to a 65 functional, pharmacologically active antibody. As a mirror image of a mirror image, the binding site of the anti-ids would be expected to be an analog of the original receptor.

The anti-id could then be used to identify and isolate peptides from banks of chemically or biologically produced banks of peptides. Selected peptides would then act as the pharmacore.

Thus, one may design drugs which have, e.g., improved BRCA2 polypeptide activity or stability or which act as inhibitors, agonists, antagonists, etc. of BRCA2 polypeptide activity. By virtue of the availability of cloned BRCA2 sequences, sufficient amounts of the BRCA2 polypeptide may be made available to perform such analytical studies as x-ray crystallography. In addition, the knowledge of the BRCA2 protein sequence provided herein will guide those employing computer modeling techniques in place of, or in addition to x-ray crystallography.

Methods of Use: Gene Therapy

According to the present invention, a method is also provided of supplying wild-type BRCA2 function to a cell which carries mutant BRCA2 alleles. Supplying such a function should suppress neoplastic growth of the recipient cells. The wild-type BRCA2 gene or a part of the gene may be introduced into the cell in a vector such that the gene remains extrachromosomal. In such a situation, the gene will be expressed by the cell from the extrachromosomal location. If a gene fragment is introduced and expressed in a cell carrying a mutant BRCA2 allele, the gene fragment should encode a part of the BRCA2 protein which is required for non-neoplastic growth of the cell. More preferred is the situation where the wild-type BRCA2 gene or a part thereof is introduced into the mutant cell in such a way that it recombines with the endogenous mutant BRCA2 gene present in the cell. Such recombination requires a double recombination event which results in the correction of the BRCA2 gene mutation. Vectors for introduction of genes both for recombination and for extrachromosomal maintenance are known in the art, and any suitable vector may be used. Methods for introducing DNA into cells such as electroporation, calcium phosphate co-precipitation and viral transduction are known in the art, and the choice of method is within the competence of the routineer. Cells transformed with the wild-type BRCA2 gene can be used as model systems to study cancer remission and drug treatments which promote such remission.

As generally discussed above, the BRCA2 gene or fragment, where applicable, may be employed in gene therapy methods in order to increase the amount of the expression products of such genes in cancer cells. Such gene therapy is particularly appropriate for use in both cancerous and pre-cancerous cells, in which the level of BRCA2 polypeptide is absent or diminished compared to normal cells. It may also be useful to increase the level of expression of a given BRCA2 gene even in those tumor cells in which the mutant gene is expressed at a "normal" level, but the gene product is not fully functional.

Gene therapy would be carried out according to generally accepted methods, for example, as described by Friedman, 1991. Cells from a patient's tumor would be first analyzed by the diagnostic methods described above, to ascertain the production of BRCA2 polypeptide in the tumor cells. A virus or plasmid vector (see further details below), containing a copy of the BRCA2 gene linked to expression control elements and capable of replicating inside the tumor cells, is prepared. Suitable vectors are known, such as disclosed in U.S. Pat. No. 5,252,479 and PCT published application WO 93/07282. The vector is then injected into the patient, either locally at the site of the tumor or systemically (in order to reach any tumor cells that may have metastasized to other sites). If the transfected gene is not permanently incorpo-

rated into the genome of each of the targeted tumor cells, the treatment may have to be repeated periodically.

Gene transfer systems known in the art may be useful in the practice of the gene therapy methods of the present invention. These include viral and nonviral transfer methods. A number of viruses have been used as gene transfer vectors, including papovaviruses, e.g., SV40 (Madzak et al., 1992), adenovirus (Berkner, 1992; Berkner et al., 1988; Gorziglia and Kapikian, 1992; Quantin et al., 1992; Rosenfeld et al., 1992; Wilkinson et al., 1992; Stratford-Pericaudet et al., 1990), vaccinia virus (Moss, 1992), adeno-associated virus (Muzyczka, 1992; Ohi et al., 1990), herpesviruses including HSV and EBV (Margolskee, 1992; Johnson et al., 1992; Fink et al., 1992; Breakfield and Geller, 1987; Freese et al., 1990), and retroviruses of avian (Brandyopadhyay and Temin, 1984; Petropoulos et al., 1992), murine (Miller, 1992; Miller et al., 1985; Sorge et al., 1984; Mann and Baltimore, 1985; Miller et al., 1988), and human origin (Shimada et al., 1991; Helseth et al., 1990; Page et al., 1990; Buchschacher and Panganiban, 1992). Most human gene therapy protocols have been based on disabled murine retroviruses.

Nonviral gene transfer methods known in the art include chemical techniques such as calcium phosphate coprecipitation (Graham and van der Eb, 1973; Pellicer et al., 1980); mechanical techniques, for example microinjection (Anderson et al., 1980; Gordon et al., 1980; Brinster et al., 1981; Constantini and Lacy, 1981); membrane fusion-mediated transfer via liposomes (Felgner et al., 1987; Wang and Huang, 1989; Kaneda et al., 1989; Stewart et al., 1992; Nabel et al., 1990; Lim et al., 1992); and direct DNA uptake and receptor-mediated DNA transfer (Wolff et al., 1990; Wu et al., 1991; Zenke et al., 1990; Wu et al., 1989b; Wolff et al., 1991; Wagner et al., 1990; Wagner et al., 1991; Cotten et al., 1990; Curiel et al., 1991a; Curiel et al., 1991b). Viral-mediated gene transfer can be combined with direct in vivo gene transfer using liposome delivery, allowing one to direct the viral vectors to the tumor cells and not into the surrounding nondividing cells. Alternatively, the retroviral vector producer cell line can be injected into tumors (Culver et al., 1992). Injection of producer cells would then provide a continuous source of vector particles. This technique has been approved for use in humans with inoperable brain tumors.

In an approach which combines biological and physical gene transfer methods, plasmid DNA of any size is combined with a polylysine-conjugated antibody specific to the adenovirus hexon protein, and the resulting complex is bound to an adenovirus vector. The trimolecular complex is then used to infect cells. The adenovirus vector permits efficient binding, internalization, and degradation of the endosome before the coupled DNA is damaged.

Liposome/DNA complexes have been shown to be capable of mediating direct in vivo gene transfer. While in standard liposome preparations the gene transfer process is nonspecific, localized in vivo uptake and expression have been reported in tumor deposits, for example, following direct in situ administration (Nabel, 1992).

Gene transfer techniques which target DNA directly to breast and ovarian tissues, e.g., epithelial cells of the breast or ovaries, is preferred. Receptor-mediated gene transfer, for example, is accomplished by the conjugation of DNA (usually in the form of covalently closed supercoiled plasmid) to a protein ligand via polylysine. Ligands are chosen on the basis of the presence of the corresponding ligand receptors on the cell surface of the target cell/tissue type. One appropriate receptor/ligand pair may include the

estrogen receptor and its ligand, estrogen (and estrogen analogues). These ligand-DNA conjugates can be injected directly into the blood if desired and are directed to the target tissue where receptor binding and internalization of the DNA-protein complex occurs. To overcome the problem of intracellular destruction of DNA, coinfection with adenovirus can be included to disrupt endosome function.

The therapy involves two steps which can be performed singly or jointly. In the first step, prepubescent females who carry a BRCA2 susceptibility allele are treated with a gene delivery vehicle such that some or all of their mammary ductal epithelial precursor cells receive at least one additional copy of a functional normal BRCA2 allele. In this step, the treated individuals have reduced risk of breast cancer to the extent that the effect of the susceptible allele has been countered by the presence of the normal allele. In the second step of a preventive therapy, predisposed young females, in particular women who have received the proposed gene therapeutic treatment, undergo hormonal therapy to mimic the effects on the breast of a full term pregnancy. Methods of Use: Peptide Therapy

Peptides which have BRCA2 activity can be supplied to cells which carry mutant or missing BRCA2 alleles. The sequence of the BRCA2 protein is disclosed in SEQ ID NO:2. Protein can be produced by expression of the cDNA sequence in bacteria, for example, using known expression vectors. Alternatively, BRCA2 polypeptide can be extracted from BRCA2-producing mammalian cells. In addition, the techniques of synthetic chemistry can be employed to synthesize BRCA2 protein. Any of such techniques can provide the preparation of the present invention which comprises the BRCA2 protein. The preparation is substantially free of other human proteins. This is most readily accomplished by synthesis in a microorganism or in vitro.

Active BRCA2 molecules can be introduced into cells by microinjection or by use of liposomes, for example. Alternatively, some active molecules may be taken up by cells, actively or by diffusion. Extracellular application of the BRCA2 gene product may be sufficient to affect tumor growth. Supply of molecules with BRCA2 activity should lead to partial reversal of the neoplastic state. Other molecules with BRCA2 activity (for example, peptides, drugs or organic compounds) may also be used to effect such a reversal. Modified polypeptides having substantially similar function are also used for peptide therapy.

Methods of Use: Transformed Hosts

Similarly, cells and animals which carry a mutant BRCA2 allele can be used as model systems to study and test for substances which have potential as therapeutic agents. The cells are typically cultured epithelial cells. These may be isolated from individuals with BRCA2 mutations, either somatic or germline. Alternatively, the cell line can be engineered to carry the mutation in the BRCA2 allele, as described above. After a test substance is applied to the cells, the neoplastically transformed phenotype of the cell is determined. Any trait of neoplastically transformed cells can be assessed, including anchorage-independent growth, tumorigenicity in nude mice, invasiveness of cells, and growth factor dependence. Assays for each of these traits are known in the art.

Animals for testing therapeutic agents can be selected after mutagenesis of whole animals or after treatment of germline cells or zygotes. Such treatments include insertion of mutant BRCA2 alleles, usually from a second animal species, as well as insertion of disrupted homologous genes.

Alternatively, the endogenous BRCA2 gene(s) of the animals may be disrupted by insertion or deletion mutation

or other genetic alterations using conventional techniques (Capecechi, 1989; Valancius and Smithies, 1991; Hasty et al., 1991; Shinkai et al., 1992; Mombaerts et al., 1992; Philpott et al., 1992; Snouwaert et al., 1992; Donehower et al., 1992). After test substances have been administered to the animals, the growth of tumors must be assessed. If the test substance prevents or suppresses the growth of tumors, then the test substance is a candidate therapeutic agent for the treatment of the cancers identified herein. These animal models provide an extremely important testing vehicle for potential therapeutic products.

The present invention is described by reference to the following Examples, which are offered by way of illustration and are not intended to limit the invention in any manner. Standard techniques well known in the art or the techniques specifically described below were utilized.

EXAMPLE 1

Ascertain and Study Kindreds Likely to Have a Chromosome 13-Linked Breast Cancer Susceptibility Locus

Extensive cancer prone kindreds were ascertained from a defined population providing a large set of extended kindreds with multiple cases of breast cancer and many relatives available to study. The large number of meioses present in these large kindreds provided the power to detect whether the BRCA2 locus was segregating, and increased the opportunity for informative recombinants to occur within the small region being investigated. This vastly improved the chances of establishing linkage to the BRCA2 region, and greatly facilitated the reduction of the BRCA2 region to a manageable size, which permits identification of the BRCA2 locus itself.

Each kindred was extended through all available connecting relatives, and to all informative first degree relatives of each proband or cancer case. For these kindreds, additional breast cancer cases and individuals with cancer at other sites of interest who also appeared in the kindreds were identified through the tumor registry linked files. All breast cancers reported in the kindred which were not confirmed in the Utah Cancer Registry were researched. Medical records or death certificates were obtained for confirmation of all cancers. Each key connecting individual and all informative individuals were invited to participate by providing a blood sample from which DNA was extracted. We also sampled spouses and relatives of deceased cases so that the genotype of the deceased cases could be inferred from the genotypes of their relatives.

Kindreds which had three or more cancer cases with inferable genotypes were selected for linkage studies to chromosome 13 markers. These included kindreds originally ascertained from the linked databases for a study of proliferative breast disease and breast cancer (Skolnick et al., 1990). The criterion for selection of these kindreds was the presence of two sisters or a mother and her daughter with breast cancer. Additionally, kindreds which have been studied since 1980 as part of our breast cancer linkage studies and kindreds ascertained from the linked databases for the presence of clusters of male and female breast cancer and self-referred kindreds with early onset breast cancer were included. These kindreds were investigated and expanded in our clinic in the manner described above.

For each sample collected in these kindreds, DNA was extracted from blood or paraffin-embedded tissue blocks using standard laboratory protocols. Genotyping in this

6,033,857

35

study was restricted to short tandem repeat (STR) markers since, in general, they have high heterozygosity and PCR methods offer rapid turnaround while using very small amounts of DNA. To aid in this effort, STR markers on chromosome 13 were developed by screening a chromosome specific cosmid library for clones which contained short tandem repeats of 2, 3 or 4, localized to the short arm in the region of the Rb tumor suppressor locus. Oligonucleotide sequences for markers not developed in our laboratory were obtained from published reports, or as part of the Breast Cancer Linkage Consortium, or from other investi-

36

relatively small region on chromosome 13 was required. Our approach was to analyze existing STR markers provided by other investigators and any newly developed markers from our laboratory in our chromosome linked kindreds. FIG. 1 shows the location of ten markers used in the genetic analysis. Table 1 gives the LOD scores for linkage for each of the 19 kindreds in our study, which reduced the region to approximately 1.5 Mb.

TABLE 1

Haplotype and Phenotype Data for the 18 Families																	
Number of Cancer Cases(1)					Posterior Probability (2)	STRs Examined											
						tdj	D13S	mb	D13S	5370-	D13S	D13S	D13S	D13S	D13S	D13S	D13S
Kindred	FBR	MBR	OV	LOD		3820	4247	260	GA9	561	171	2C	A6C	310	267		
107*	22	3	2	5.06	1.00	8	28	4	10	8	3	2	6	4	12		
8001	0	3	0	n.d.	0.90	8	30	6	10	7	10	5	5	5	4		
8004	1	2	0	n.d.	0.90	9	11	4	4	7	8	6	8	4	12		
2044*	8	1	4	2.13	1.00	9	12	10	7	5	9	6	5	4	8		
2043*	2	1	1	0.86	0.98	6	30	3	12	7	10	5	8	4	12		
2018	3	1	0	n.d.	0.90	9	12	7	3	8	3	6	6	5	8		
937	3	1	0	n.d.	0.90	8	10	4	—	—	8	10	6	7	7		
1018*	9	1	0	2.47	1.00	6	17	8	10	5	8	2	5	4	8		
2328	11	1	0	0.42	0.96	9	10	3	10	5	8	5	5	7	12		
2263	2	1	0	n.d.	0.90	9	28	8	—	8	4	—	—	7	12		
8002	2	1	0	n.d.	0.90	3	29	7	10	5	8	5	5	5	8		
8003	2	1	0	n.d.	0.90	4	12	6	10	6	3	4	5	4	8		
2367	6	0	1	0.40	0.85	6	28	7	10	12	3	7	5	5	4		
2388	3	0	1	0.92	0.95	8	16	7	12	4	10	4	5	5	12		
2027*	4	0	0	0.39	0.85	4	11	3	10	7	10	5	6	7	12		
4328	4	0	0	0.44	0.87	9	10	8	4	8	3	7	8	5	12		
2355	3	0	0	0.36	0.84	9	10	6	4	6	3	7	3	5	8		
2327	11	0	0	1.92	0.99	3	12	2	9	5	10	5	5	3	4		
1019	2	2	0														

*Families reported in Wooster et al. (1994).
n.d. = not determined
(1)Excludes cases known to be sporadic (i.e., do not share the BRCA2 haplotype segregating in the family).
FBR = female breast cancer under 60 years.
MBR = male breast cancer
OV = ovarian cancer
(2) Posterior probability assumes that, a priori, 90% of families with male breast and early onset female breast cancers that are unlinked to BRCA1 are due to BRCA2, and 70% of female breast cancer families unlinked to BRCA1 are due to BRCA1.

gators. All genotyping films were scored blindly with a standard lane marker used to maintain consistent coding of alleles. Key samples underwent duplicate typing for all relevant markers.

LOD scores for each kindred were calculated for two recombination fraction values, 0.001 and 0.1. (For calculation of LOD scores, see Ott 1985). Likelihoods were computed under the model derived by Claus et al., 1991, which assumes an estimated gene frequency of 0.003, a lifetime risk in female gene carriers of about 0.80, and population based age-specific risks for breast cancer in non-gene carriers. Allele frequencies for the markers used for the LOD score calculations were calculated from our own laboratory typings of unrelated individuals in the CEPH panel (White and Lalouel, 1988).

Kindred 107 is the largest chromosome 13-linked breast cancer family reported to date by any group. The evidence of linkage to chromosome 13 for this family is overwhelming. In smaller kindreds, sporadic cancers greatly confound the analysis of linkage and the correct identification of key recombinants.

In order to improve the characterization of our recombinants and define closer flanking markers, a dense map of this

Table 1 also gives the posterior probability of a kindred having a BRCA2 mutation based on LOD scores and prior probabilities. Four of these markers (D13S171, D13S260, D13S310 and D13S267) were previously known. The other six markers were found as part of our search for BRCA2. We were able to reduce the region to 1.5 megabases based on a recombinant in Kindred 107 with marker tdj3820 at the left boundary, and a second recombinant in Kindred 2043 with marker YS-G-B10T at the right boundary (see FIG. 1) which is at approximately the same location as AC6 and D13S310. Furthermore, a homozygous deletion was found in a pancreatic tumor cell line in the BRCA2 region which may have been driven by BRCA2 itself; this deletion is referred to as the Schutte/Kern deletion in FIG. 1 (Schutte et al., 1995). The Schutte/Kern contig in FIG. 1 refers to these authors' physical map which covers the deletion.

EXAMPLE 2

Development of Genetic and Physical Resources in the Region of Interest

To increase the number of highly polymorphic loci in the BRCA2 region, we developed a number of STR markers in

6,033,857

37

our laboratory from P1s, BACs and YACs which physically map to the region. These markers allowed us to further refine the region (see Table 1 and the discussion above).

STSs in the desired region were used to identify YACs which contained them. These YACs were then used to identify subclones in P1s or BACs. These subclones were then screened for the presence of a short tandem repeats. Clones with a strong signal were selected preferentially, since they were more likely to represent repeats which have a large number of repeats and/or are of near-perfect fidelity to the pattern. Both of these characteristics are known to increase the probability of polymorphism (Weber et al., 1990). These clones were sequenced directly from the vector to locate the repeat. We obtained a unique sequence on one side of the short tandem repeat by using one of a set of possible primers complementary to the end of the repeat. Based on this unique sequence, a primer was made to sequence back across the repeat in the other direction, yielding a unique sequence for design of a second primer flanking it. STRs were then screened for polymorphism on a small group of unrelated individuals and tested against the hybrid panel to confirm their physical localization. New markers which satisfied these criteria were then typed in a set of unrelated individuals from Utah to obtain allele frequencies appropriate for the study of this population. Many of the other markers reported in this study were also tested in unrelated individuals to obtain similarly appropriate allele frequencies.

Using the procedure described above, novel STRs were found from these YACs which were both polymorphic and localized to the BRCA2 region. FIG. 1 shows a schematic map of STSs, P1s, BACs and YACs in the BRCA2 region.

EXAMPLE 3

Identification of Candidate cDNA Clones for the BRCA2 Locus by Genomic Analysis of the Contig Region

1. General Methods

Complete screen of the plausible region. The first method to identify candidate cDNAs, although labor intensive, used known techniques. The method comprised the screening of P1 and BAC clones in the contig to identify putative coding sequences. The clones containing putative coding sequences were then used as probes on filters of cDNA libraries to identify candidate cDNA clones for future analysis. The clones were screened for putative coding sequences by either of two methods.

The P1 clones to be analyzed were digested with a restriction enzyme to release the human DNA from the vector DNA. The DNA was separated on a 14 cm, 0.5% agarose gel run overnight at 20 volts for 16 hours. The human DNA bands were cut out of the gel and electroeluted from the gel wedge at 100 volts for at least two hours in 0.5x Tris Acetate buffer (Maniatis et al., 1982).

The eluted Not I digested DNA (~15 kb to 25 kb) was then digested with EcoRI restriction enzyme to give smaller fragments (~0.5 kb to 5.0 kb) which melt apart more easily for the next step of labeling the DNA with radionucleotides. The DNA fragments were labeled by means of the hexamer random prime labeling method (Boehringer-Mannheim, Cat. #1004760). The labeled DNA was spermine precipitated (add 100 µl TE, 5 µl 0.1 M spermine, and 5 µl of 10 mg/ml salmon sperm DNA) to remove unincorporated radionucleotides. The labeled DNA was then resuspended in 100 µl TE, 0.5 M NaCl at 65° C. for 5 minutes and then blocked with Human C_αt-1 DNA for 2-4 hrs. as per the manufacturer's

38

instructions (Gibco/BRL, Cat. #5279SA). The C_αt-1 blocked probe was incubated on the filters in the blocking solution overnight at 42° C. The filters were washed for 30 minutes at room temperature in 2xSSC, 0.1% SDS, and then in the same buffer for 30 minutes at 55° C. The filters were then exposed 1 to 3 days at -70° C. to Kodak XAR-5 film with an intensifying screen. Thus, the blots were hybridized with either the pool of EcoRI fragments from the insert, or each of the fragments individually.

The human DNA from clones in the region was isolated as whole insert or as EcoRI fragments and labeled as described above. The labeled DNA was used to screen filters of various cDNA libraries under the same conditions described above except that the cDNA filters undergo a more stringent wash of 0.1xSSC, 0.1% SDS at 65° C. for 30 minutes twice.

Most of the cDNA libraries used to date in our studies (libraries from normal breast tissue, breast tissue from a woman in her eighth month of pregnancy and a breast malignancy) were prepared at Clontech, Inc. The cDNA library generated from breast tissue of an 8 month pregnant woman is available from Clontech (Cat. #HL1037a) in the Lambda gt-10 vector, and is grown in C600Hf1 bacterial host cells. Normal breast tissue and malignant breast tissue samples were isolated from a 37 year old Caucasian female and one-gram of each tissue was sent to Clontech for mRNA processing and cDNA library construction. The latter two libraries were generated using both random and oligo-dT priming, with size selection of the final products which were then cloned into the Lambda Zap II vector, and grown in XL1-blue strain of bacteria as described by the manufacturer. Additional tissue-specific cDNA libraries include human fetal brain (Stratagene, Cat. 936206), human testis (Clontech Cat. HL3024), human thymus (Clontech Cat. HL1127n), human brain (Clontech Cat. HL11810), human placenta (Clontech Cat 1075b), and human skeletal muscle (Clontech Cat. HL1124b).

The cDNA libraries were plated with their host cells on NZCYM plates, and filter lifts are made in duplicate from each plate as per Maniatis et al. (1982). Insert (human) DNA from the candidate genomic clones was purified and radioactively labeled to high specific activity. The radioactive DNA was then hybridized to the cDNA filters to identify those cDNAs which correspond to genes located within the candidate cosmid clone. cDNAs identified by this method were picked, replated, and screened again with the labeled clone insert or its derived EcoRI fragment DNA to verify their positive status. Clones that were positive after this second round of screening were then grown up and their DNA purified for Southern blot analysis and sequencing. Clones were either purified as plasmid through in vivo excision of the plasmid from the Lambda vector as described in the protocols from the manufacturers, or isolated from the Lambda vector as a restriction fragment and subcloned into plasmid vector.

The Southern blot analysis was performed in duplicate, one using the original genomic insert DNA as a probe to verify that cDNA insert contains hybridizing sequences. The second blot was hybridized with cDNA insert DNA from the largest cDNA clone to identify which clones represent the same gene. All cDNAs which hybridize with the genomic clone and are unique were sequenced and the DNA analyzed to determine if the sequences represent known or unique genes.

All cDNA clones which appear to be unique were further analyzed as candidate BRCA2 loci. Specifically, the clones are hybridized to Northern blots to look for breast specific

expression and differential expression in normal versus breast tumor RNAs. They are also analyzed by PCR on =clones in the BRCA2 region to verify their location. To map the extent of the locus, full length cDNAs are isolated and their sequences used as PCR probes on the YACs and the clones surrounding and including the original identifying clones. Intron-exon boundaries are then further defined through sequence analysis.

We have screened the normal breast, 8 month pregnant breast and fetal brain cDNA libraries with Eco RI fragments from cosmid BAC and P1 clones in the region. Potential BRCA2 cDNA clones were identified among the three libraries. Clones were picked, replated, and screened again with the original probe to verify that they were positive.

Analysis of hybrid-selected cDNA. cDNA fragments obtained from direct selection were checked by Southern blot hybridization against the probe DNA to verify that they originated from the contig. Those that passed this test were sequenced in their entirety. The set of DNA sequences obtained in this way were then checked against each other to find independent clones that overlapped.

The direct selection of cDNA method (Lovett et al., 1991; Futreal, 1993) is utilized with P1 and BAC DNA as the probe. The probe DNA is digested with a blunt cutting restriction enzyme such as HaeIII. Double-stranded adapters are then ligated onto the DNA and serve as binding sites for primers in subsequent PCR amplification reactions using biotinylated primers. Target cDNA is (generated from mRNA derived from tissue samples, e.g., breast tissue, by synthesis of either random primed or oligo(dT) primed first strand, followed by second strand synthesis. The cDNA ends are rendered blunt and ligated onto double-stranded adapters. These adapters serve as amplification sites for PCR. The target and probe sequences are denatured and mixed with human C₆-1 DNA to block repetitive sequences. Solution hybridization is carried out to high C₆-1/2 values to ensure hybridization of rare target cDNA molecules. The annealed material is then captured on avidin beads, washed at high stringency and the retained cDNAs are eluted and amplified by PCR. The selected cDNA is subjected to further rounds of enrichment before cloning into a plasmid vector for analysis.

HTF island analysis. A method for identifying cosmids to use as probes on the cDNA libraries was HTF island analysis. HTF islands are segments of DNA which contain a very high frequency of unmethylated CpG dinucleotides (Tonolio et al., 1990) and are revealed by the clustering of restriction sites of enzymes whose recognition sequences include CpG dinucleotides. Enzymes known to be useful in HTF-island analysis are AscI, NotI, BssHII, EagI, SacII, NaeI, NarI, SmaI, and MluI (Anand, 1992).

Analysis of candidate clones. One or more of the candidate genes generated from above were sequenced and the information used for identification and classification of each expressed gene. The DNA sequences were compared to known genes by nucleotide sequence comparisons and by translation in all frames followed by a comparison with known amino acid sequences. This was accomplished using Genetic Data Environment (GDE) version 2.2 software and the Basic Local Alignment Search Tool (Blast) series of client/server software packages (e.g., BLASTN 1.3.13MP), for sequence comparison against both local and remote sequence databases (e.g., GenBank), running on Sun SPARC workstations. Sequences reconstructed from collections of cDNA clones identified with the cosmids and P1s have been generated. All candidate genes that represented new sequences were analyzed further to test their candidacy for the putative BRCA2 locus.

Mutation screening. To screen for mutations in the affected pedigrees, two different approaches were followed. First, genomic DNA isolated from family members known to carry the susceptibility allele of BRCA2 was used as a template for amplification of candidate gene sequences by PCR. If the PCR primers flank or overlap an intron/exon boundary, the amplified fragment will be larger than predicted from the cDNA sequence or will not be present in the amplified mixture. By a combination of such amplification experiments and sequencing of P1 or BAC clones using the set of designed primers it is possible to establish the intron/exon structure and ultimately obtain the DNA sequences of genomic DNA from the kindreds.

A second approach that is much more rapid if the intron/exon structure of the candidate gene is complex involves sequencing fragments amplified from cDNA synthesized from lymphocyte mRNA extracted from pedigree blood which was used as a substrate for PCR amplification using the set of designed primers. If the candidate gene is expressed to a significant extent in lymphocytes, such experiments usually produce amplified fragments that can be sequenced directly without knowledge of intron/exon junctions.

The products of such sequencing reactions were analyzed by gel electrophoresis to determine positions in the sequence that contain either mutations such as deletions or insertions, or base pair substitutions that cause amino acid changes or other detrimental effects.

Any sequence within the BRCA2 region that is expressed in breast is considered to be a =candidate gene for BRCA2. Compelling evidence that a given candidate gene corresponds to BRCA2 comes from a demonstration that kindred families contain defective alleles of the candidate.

2. Specific Methods

Hybrid selection. Two distinct methods of hybrid selection were used in this work.

Method 1: cDNA preparation and selection. Randomly primed cDNA was prepared from poly (A)⁺RNA of mammary gland, ovary testis, fetal brain and placenta tissues and from total RNA of the cell line Caco-2 (ATCC HTB 37). cDNAs were homopolymer tailed and then hybrid selected for two consecutive rounds of hybridization to immobilized P1 or BAC DNA as described previously. (Parimoo et al., 1991; Rommens et al., 1994). Groups of two to four overlapping P1 and/or BAC clones were used in individual selection experiments. Hybridizing cDNA was collected, passed over a G50 Fine Sephadex column and amplified using tailed primers. The products were then digested with EcoRI, size selected on agarose gels, and ligated into pBluescript (Stratagene) that had been digested with EcoRI and treated with calf alkaline phosphatase (Boehringer Mannheim). Ligation products were transformed into competent DH5a *E. coli* cells (Life Technologies, Inc.).

Characterization of Retrieved cDNAs. 200 to 300 individual colonies from each ligation (from each 250 kbases of genomic DNA) were picked and gridded into microtiter plates for ordering and storage. Cultures were replica transferred onto Hybond N membranes (Amersham) supported by LB agar with ampicillin. Colonies were allowed to propagate and were subsequently lysed with standard procedures. Initial analysis of the cDNA clones involved a prescreen for ribosomal sequences and subsequent cross screenings for detection of overlap and redundancy.

Approximately 10–25% of the clones were eliminated as they hybridized strongly with radiolabeled cDNA obtained from total RNA. Plasmids from 25 to 50 clones from each selection experiment that did not hybridize in prescreening

6,033,857

41

were isolated for further analysis. The retrieved cDNA fragments were verified to originate from individual starting genomic clones by hybridization to restriction digests of DNAs of the starting clones, of a hamster hybrid cell line (GM10898A) that contains chromosome 13 as its only human material and to human genomic DNA. The clones were tentatively assigned into groups based on the overlapping or non-overlapping intervals of the genomic clones. Of the clones tested, approximately 85% mapped appropriately to the starting clones.

Method 2 (Lovett et al., 1991): cDNA Preparation. Poly (A) enriched RNA from human mammary gland, brain, lymphocyte and stomach were reverse-transcribed using the tailed random primer XN₁₂

[5'-(NH₂)-GTAGTGCAAGGCTCGAGAACNNNNNNNNN] (SEQ ID NO:3)

and Superscript II reverse transcriptase (Gibco BRL). After second strand synthesis and end polishing, the ds cDNA was purified on Sepharose CL-4B columns (Pharmacia). cDNAs were "anchored" by ligation of a double-stranded oligo RP

[5'-(NH₂)-TGAGTAGAATTCTAACGCCGTCATTGTTC (SEQ ID NO:4)

annealed to

5'-GAACAATGACGCCGTTAGAATTCTACTCA-(NH₂) (SEQ ID NO:5)]

to their 5' ends (5' relative to mRNA) using T4 DNA ligase. Anchored ds cDNA was then repurified on Sepharose CL-4B columns.

Selection. cDNAs from mammary gland, brain, lymphocyte and stomach tissues were first amplified using a nested version of RP

(RPA: 5'-TGAGTAGAATTCTAACGCCGTCAT' (SEQ ID NO:6)

and

XPCR [5'-(PO₄)-GTAGTGCAAGGCTCGAGAAC (SEQ ID NO:7)]

and purified by fractionation on Sepharose CL-4B. Selection probes were prepared from purified PIs, BACs or PACs by digestion with HinfI and Exonuclease III. The single-stranded probe was photolabelled with photobiotin (Gibco BRL) according to the manufacturer's recommendations. Probe, cDNA and Cot-1 DNA were hybridized in 2.4M TEA-CL, 10 mM NaPO₄, 1 mM EDTA. Hybridized cDNAs were captured on streptavidin-paramagnetic particles (Dynal), eluted, reamplified with a further nested version of RP

[RPB: 5'-(PO₄)-TGAGTAGAATTCTAACGCCGTCATTG (SEQ ID NO:8)]

and XPCR, and size-selected on Sepharose CL-6B. The selected, amplified cDNA was hybridized with an additional aliquot of probe and C₀t-1 DNA. Captured and eluted products were amplified again with RPB and XPCR, size-selected by gel electrophoresis and cloned into dephosphorylated HincII cut pUC18. Ligation products were transformed into XL2-Blue ultra-competent cells (Stratagene).

Analysis. Approximately 192 colonies for each single-probe selection experiment were amplified by colony PCR using vector primers and blotted in duplicate onto Zeta Probe nylon filters (Bio-Rad). The filters were hybridized

42

using standard procedures with either random primed C₀t-1 DNA or probe DNA (P1, BAC or PAC). Probe-positive, C₀t-1 negative clones were sequenced in both directions using vector primers on an ABI 377 sequencer.

Exon Trapping. Exon amplification was performed using a minimally overlapping set of BACs, PIs and PACs in order to isolate a number of gene sequences from the BRCA2 candidate region. Pools of genomic clones were assembled, containing from 100-300 kb of DNA in the form of 1-3 overlapping genomic clones. Genomic clones were digested with PstI or BamHI+BglII and ligated into PstI or BamHI sites of the pSPL3 splicing vector. The exon amplification technique was performed (Church et al., 1993) and the end products were cloned in the pAMP1 plasmid from the Uracil DNA Glycosylase cloning system (BRL). Approximately 6000 clones were picked, propagated in 96 well plates, stamped onto filters, and analyzed for the presence of vector and repeat sequences by hybridization. Each clone insert was PCR amplified and tested for redundancy, localization and human specificity by hybridization to grids of exons and dot blots of the parent genomic DNA. Unique candidate exons were sequenced, searched against the databases, and used for hybridization to cDNA libraries.

5' RACE. The 5' end of BRCA2 was identified by a modified RACE protocol called biotin capture RACE. Poly (A) enriched RNA from human mammary gland and thymus was reverse-transcribed using the tailed random primer XN₁₂

[5'-(NH₂)-GTAGTGCAAGGCTCGAGAACNNN (SEQ ID NO:3)]

and Superscript II reverse transcriptase (Gibco BRL). The RNA strand was hydrolyzed in NaOH and first strand cDNA purified by fractionation on Sepharose CL-4B (Pharmacia). First strand cDNAs were "anchored" by ligation of a double-stranded oligo with a 7 bp random 5' overhang [ds UCA: 5'-CCTTCACACGCGTATCGATTAGTCACNNNNNNNN-(NH₂) (SEQ ID NO:9) annealed to 5'-(PO₄)-GTGACTAATCGATACGCGTGTGAAGGTGC (SEQ ID NO:10)] to their 3' ends using T4 DNA ligase. After ligation, the anchored cDNA was repurified by fractionation on Sepharose CL-4B. The 5' end of BRCA2 was amplified using a biotinylated reverse primer [5'-(B)-TTGAAGAACAACAGGACTTTTCTACTA] (SEQ ID NO:11) and a nested version of UCA [UCP.A: 5'-CACCTTCACACGCGTATCG (SEQ ID NO:12)]. PCR products were fractionated, on an agarose gel, gel purified, and captured on streptavidin-paramagnetic particles (Dynal). Captured cDNA was reamplified using a nested reverse primer [5'-GTTCGTAATTGTTGTTTATGTTTCAG] (SEQ ID NO:13) and a further nested version of UCA [UCP.B: 5'-CCTTCACACGCGTATCGATTAG] (SEQ ID NO:14)]. This PCR reaction gave a single sharp band on an agarose gel; the DNA was gel purified and sequenced in both directions on an ABI 377 sequencer.

cDNA Clones. Human cDNA libraries were screened with ³²P-labeled hybrid selected or exon trapped clones. Phage eluted from tertiary plaques were PCR amplified with vector-specific primers and then sequenced on an ABI 377 sequencer.

Northern Blots. Multiple Tissue Northern (MTN) filters, which are loaded with 2 µg per lane of poly(A)+RNA derived from a number of human tissues, were purchased from Clontech. ³²P-random-primer labeled probes corresponding to retrieved cDNAs GT 713 (BRCA2 exons 3-7), k wCPF1B8.1 (3' end of exon 11 into exon 20), and glyceraldehyde-3-phosphate dehydrogenase (GAPDH) were

used to probe the filters. Prehybridizations were at 42° C. in 50% formamide, 5×SSPE, 1% SDS, 5×Denhardt's mixture, 0.2 mg/ml denatured salmon testis DNA and 2 μg/ml poly (A). Hybridizations were in the same solution with the addition of dextran sulfate to 4% and probe. Stringency washes were in 0.1×SSC/0.1% SDS at

RT-PCR Analysis. Ten μg of total RNA extracted from five human breast cancer cell lines (ZR-75-1, T-47D, MDA-MB-231, MDA-MB468 and BT-20) and three human prostate cancer cell lines (LNCaP, DU145 and PC-3) (RNAs provided by Dr. Claude Labrie, CHUL Research Center) were reverse transcribed using the primer mH20-D105#RA

[5'-TTTGGATCATTTTCACACTGTC] (SEQ ID NO:15)]

and Superscript II reverse transcriptase (Gibco BRL). Thereafter, the single strand cDNAs were amplified using the primers CG0269FB:

[5'-GTGCTCATAGTCAGAAATGAAG] (SEQ ID NO:16)]

and mH20-1D05#RA (this is the primer pair that was used to island hop from the exon 7/8 junction into exon 11; the PCR product is about 1.55 kb). PCR products were fractionated on a 1.2% agarose gel.

PCR Amplification and Mutation Screening. All 26 coding exons of BRCA2 and their associated splice sites were amplified from genomic DNA as described (Kamb et al.,

1994b). The DNA sequences of the primers, some of which lie in flanking intron sequence, used for amplification and sequencing appear in Table 2. Some of the exons (2 through 10, 11-5, 11-6, 11-7 and 23 through 27) were amplified by a simple one-step method. The PCR conditions for those exons were: single denaturing step of 95° C. (1 min.); 40 cycles of 96° C. (6 sec.), T_{ann}=55° C. (15 sec.), 72° C. (1 min.). Other exons (11-22) required nested reamplification after the primary PCR reaction. In these cases, the initial amplification was carried out with the primers in the first two columns of Table 2 for 19 cycles as described above. Nested reamplification for these exons was carried out for 28 or 32 cycles at the same conditions with the primers appearing in the third column of Table 2. The buffer conditions were as described (Kamb et al., 1994b). The products were purified from 0.8% agarose gels using Qiaex beads (Qiagen). The purified products were analyzed by cycle sequencing with α-P³²dATP with Ampli-Cycle™ Sequencing Kit (Perkin Elmer, Branchburg, N.J.). The reaction products were fractionated on 6% polyacrylamide gels. All (A) reactions were loaded adjacent each other, followed by the (C) reactions, etc. Detection of polymorphisms was carried out visually and confirmed on the other strand.

TABLE 2

Primers for Amplifying BRCA2 Exons			
EXON	FORWARD PRIMER	REVERSE PRIMER	NESTED PRIMER
2	TGTTCCCATCTCACAGTAAG ^{*(17)}	GTACTGGGTTTTTAGCAAGCA ^{*(18)}	
3	GGTTAAAACTAAGGTGGGA ^{*(19)}	ATTTGCCAGCATGACACA ^{*(20)}	
4	TTTCCAGTATAGAGGAGA ^{*(21)}	GTAGGAAAATGTTTCATTTAA ^{*(22)}	
5	ATCTAAAGTAGTATTCACAACA ^{*(23)}	GGGGGTAAAAAAGGGGAA ^{*(24)}	
6	GAGATAAGTCAGGTATGATT ^{*(25)}	AATTGCCTGTATGAGGCAGA ^{*(26)}	
7	GGCAATTCAGTAAACGTTAA ^{*(27)}	ATTGTCAGTTACTAACACAC ^{*(28)}	
8	GTGTCAITGTAATCAAAATAGT ^{*(29)}	CAGGTTTAGAGACTTTCTC ^{*(30)}	
9	GGACCTAGGTTGATTGCA ^{*(31)}	GTCAAGAAAGGTAAGGTAA ^{*(32)}	
10-1	CTATGAGAAAGGTTGTGAG ^{*(33)}	CCTAGTCTTGCTAGTTCTT ^{*(34)}	
10-2	AACAGTTGTAGATACCTCTGAA ^{*(35)}	GACTTTTTGTATACCCTGAAATG ^{*(36)}	
10-3	CAGCAITCTGAATCTCATACAG ^{*(37)}	CATGTATACAGATGATGCCTAAG ^{*(38)}	
11-1	AACTTAGTGAAAAATATTTAGTGA ^{*(39)}	ATACATCTTGATTTCTTTCCAT ^{*(40)}	TTTAGTGAATGTGATTGATGGT ^{*(41)}
11-2	AGAACAACCTTTGCTCTTAA ^{*(42)}	TIAGATTIGTGTTTTGGTTGAA ^{*(43)}	TAGCTCTTTTGGGACAATTC ^{*(44)}
11-3	ATGAAAAAGAATCAAGATGTAT ^{*(45)}	CCTAATGTTATGTTCTAGAGAG ^{*(46)}	GCTACCTCCAAAACCTGTGA ^{*(47)}
11-4	GTGTAAAGCAGCATATAAAAT ^{*(48)}	CTTGCTGCTGTCTACCTG ^{*(49)}	AGTGGTCTTAAAGATAGTCAI ^{*(50)}
11-5	CCATAATTTAACACCTAGCCA ^{*(51)}	CCAAAAAGTTAAATCTGACA ^{*(52)}	
	GGCTTTTATCTGCTCATGGC ^{*(53)}	CCTCTGCAGAAAGTTTCTCAC ^{*(54)}	
11-6	AACGGAATGCTATTTACTGA ^{*(55)}	AGTACCTTGCTCTTTTTCATC ^{*(56)}	
11-7	CAGCTAGCGGGAAAAAAGTTA ^{*(57)}	TTTCGGAGAGATGATTTTGTG ^{*(58)}	
11-8	GCCTTAGCTTTTACACAA ^{*(59)}	TTTTTGATTATCTCGTTG ^{*(60)}	TTATCTCGTTGTTTCTCTTA ^{*(61)}
11-9	CCATTAATTTGTCCATATCTA ^{*(62)}	GACGTAGGTGAATAGTGAAGA ^{*(63)}	TCAAATTCCTCTAACACTCC ^{*(64)}
11-10	GAAGATAGTACCAAGCAAGTC ^{*(65)}	TGAGACTTTGGTTCTCAATAC ^{*(66)}	AGTAACGAACATTCAGACCAG ^{*(67)}
11-11	GTCTTCACTATTCACCTACG ^{*(68)}	CCCCAAACTGACTACACAA ^{*(69)}	AGCATACCAAGTCTACTGAAT ^{*(70)}
12	ACTCTTTCAAACATTAGGTCA ^{*(71)}	TTGGAGAGGCAGGTGGAT ^{*(72)}	CTATAGAGGGAGAACAGAT ^{*(73)}
13	TTTATGCTGATTTCTGTTGTAT ^{*(74)}	ATAAAACGGGAAGTGTTAACT ^{*(75)}	CTGTGAGTTATTTGGTGCAI ^{*(76)}
14	GAATACAAAACAGTTACCAGA ^{*(77)}	CACCACCAAGGGGGAAA ^{*(78)}	AAATGAGGGTCTGCAACAAA ^{*(79)}
15	GTCCGACCAAGACTTGAG ^{*(80)}	AGCCATTTGTAGGATCTAG ^{*(81)}	CTACTAGACGGCGGAG ^{*(82)}
16	ATGTTTTTGTAGTGAAGATTCT ^{*(83)}	TAGTTCGAGAGACAGTTAAG ^{*(84)}	CAGTTTTGGTTTGTATAATTG ^{*(85)}
17	CAGAGAATAGTTGTAGTTGTT ^{*(86)}	AACCTTAAACCACTACTGCC ^{*(87)}	TTCAATATCATCTATGTGG ^{*(88)}
18	TTTTATCTCAGTTATTCAGTG ^{*(89)}	GAAATTGAGCATCCTTAGTAA ^{*(90)}	AATCTAGAGTCACACTTCC ^{*(91)}
19	ATATTTTAAAGGCAGTTCTAGA ^{*(92)}	TTACACACACCAAAAAAGTCA ^{*(93)}	TGAAAACTCTTTATGATATCTGT ^{*(94)}
20	TGAATGTTATATATGTGACTTTT ^{*(95)}	CTTGTTGCTATCTTTGTCTA ^{*(96)}	CCCTAGATACTAAAAAATAAG ^{*(97)}
21	CTTTTAGCAGTTATATAGTTTC ^{*(98)}	GCCAGAGAGTCTAAAAACAG ^{*(99)}	CTTTGGGTGTTTTATGCTTG ^{*(100)}
22	TTTGTGTATTTGTCTCTGTTA ^{*(101)}	ATTTGTGTTAGGTCAITTTT ^{*(102)}	GTTCGATTGCTTTTATTC ^{*(103)}

TABLE 2-continued

Primers for Amplifying BRCA2 Exons			
EXON	FORWARD PRIMER	REVERSE PRIMER	NESTED PRIMER
23	ATCACTTCTTCCATTGCATC*(104)	CCGTGGCTGGTAAATCTG*(105)	
24	CTGGTAGCTCCAACATAATC*(106)	ACCGGTACAAACCTTTCAITG*(107)	
25	CTAITTTGATTGCTTTTATTATT*(108)	GCTAITTCCTTGATCTGGAC*(109)	
26	TTGGAACATAAATATGTGGG*(110)	ACTTACAGGAGCCACATAAC*(111)	
27	CTACATTAATTATGATAGGCTNCG***(112)	GTACTAATGTGTGGTTTGAAA***(113)	TCAATGCAAGTTCCTCGTCAGC*(114)

Primers with an "*" were used for sequencing.
Primers without an "*" were replaced by the internal nested primer for both the second round of PCR and sequencing.
For large exons requiring internal sequencing primers, primers with an "***" were used to amplify the exon
Number in parenthesis refers to the SEQ ID NO: for each primer.

EXAMPLE 4

Identification of BRCA2

Assembly of the full-length BRCA2 sequence. The full-length sequence of BRCA2 was assembled by combination of several smaller sequences obtained from hybrid selection, exon trapping, cDNA library screening, genomic sequencing, and PCR experiments using cDNA as template for amplification (i.e., "island hopping") (FIG. 2). The extreme 5' end of the mRNA including the predicted translational start site was identified by a modified 5'RACE protocol (Stone et al., 1995). The first nucleotide in the sequence (nucleotide 1) is a non-template G, an indication that the mRNA cap is contained in the sequence. One of the exons (exon 11) located on the interior of the BRCA2 cDNA is nearly 5 kb. A portion of exon 11 was identified by analysis of roughly 900 kb of genomic sequence in the public domain ([ftp://genome.wustl.edu/pub/gsc1/brca](http://genome.wustl.edu/pub/gsc1/brca)). This genomic sequence was condensed with genomic sequence determined by us into a set of 160 sequence contigs. When the condensed genomic sequence was scanned for open reading frames (ORFs), a contiguous stretch of nearly 5 kb was identified that was spanned by long ORFs. This sequence was linked together by island hopping experiments with two previously identified candidate gene fragments. The current composite BRCA2 cDNA sequence consists of 11,385 bp, but does not include the polyadenylation signal or poly(A) tail. This cDNA sequence is set forth in SEQ ID NO:1 and FIG. 3.

Structure of the BRCA2 gene and BRCA2 polypeptide. Conceptual translation of the cDNA revealed an ORF that began at nucleotide 229 and encoded a predicted protein of 3418 amino acids. The peptide bears no discernible similarity to other proteins apart from sequence composition. There is no signal sequence at the amino terminus, and no obvious membrane-spanning regions. Like BRCA1, the BRCA2 protein is highly charged. Roughly one quarter of the residues are acidic or basic.

The BRCA2 gene structure was determined by comparison of cDNA and genomic sequences. BRCA2 is composed of 27 exons distributed over roughly 70 kb of genomic DNA.

A CpG-rich region at the 5' end of BRCA2 extending upstream suggests the presence of regulatory signals often associated with CpG "islands." Based on Southern blot experiments, BRCA2 appears to be unique, with no close homologs in the human genome.

Expression studies of BRCA2. Hybridization of labeled cDNA to human multiple tissue Northern filters revealed an 11–12 kb transcript that was detectable in testis only. The size of the this transcript suggests that little of the BRCA2 mRNA sequence is missing from our composite cDNA. Because the Northern filters did not include mammary gland RNA, RT-PCR experiments using a BRCA2 cDNA amplicon were performed on five breast and three prostate cancer cell line RNAs. All of the lines produced positive signals. In addition, PCR of a BRCA2 amplicon (1-BrCGO26→5kb) and 5' RACE were used to compare mammary gland and thymus cDNA as templates for amplification. In both cases, the product amplified more efficiently from breast than from thymus.

Germline mutations in BRCA2. Individuals from eighteen putative BRCA2 kindreds were screened for BRCA2 germline mutations by DNA sequence analysis (Wooster et al., 1994). Twelve kindreds have at least one case of male breast cancer, four have two or more cases; and, four include at least one individual affected with ovarian cancer who shares the linked BRCA2 haplotype. Each of the 18 kindreds has a posterior probability of harboring a BRCA2 mutation of at least 69%, and nine kindreds have posterior probabilities greater than 90%. Based on these combined probabilities, 16 of 18 kindreds are expected to segregate BRCA2 mutations. The entire coding sequence and associated splice junctions were screened for mutations in multiple individuals from nine kindreds using either cDNA or genomic DNA (Table 3). Individuals from the remaining nine kindreds were screened for mutations using only genomic DNA. These latter screening experiments encompassed 99% of the coding sequence (all exons excluding exon 15) and all but two of the splice junctions.

TABLE 3

Family	Set of Families Screened for BRCA2 Mutations						BRCA2 Mutation	Exon	Codon	Effect
	FBC	FBC <50 yrs	Ov	MBC	LOD	Prior Probability				
UT-107 ¹	20	18	2	3	5.06	1.00	277 delAC	2	17	termination codon at 29
UT-1018 ¹	11	9	0	1	2.47	1.00	982 del4	9	252	termination codon at 275
UT-2044 ¹	8	6	4	1	2.13	1.00	4706 del4	11	1493	termination codon at 1502
UT-2367 ¹	6	5	1	0	2.09	0.99	IR			
UT-2327 ¹	13	6	0	0	1.92	0.99	ND			
UT-2388 ¹	3	3	1	0	0.92	0.92	ND			
UT-2328 ¹	10	4	0	1	0.21	0.87	ND			
UT-4328 ¹	4	3	0	0	0.18	0.69	ND			
MI-1016 ¹	4	2	0	1	0.04	0.81	ND			
CU-20 ²	4	3	2	2	1.09	1.00	8525 delC	18	2766	termination codon at 2776
CU-159 ²	8	4	0	0	0.99	0.94	9254 del 5	23	3009	termination codon at 3015
UT-2043 ²	2	2	1	1	0.86	0.97	4075 delGT	11	1283	termination codon at 1285
IC-2204 ²	3	1	0	4	0.51	0.98	999 del5	9	257	termination codon at 273
MS-075 ²	4	1	0	1	0.50	0.93	6174 delT	11	1982	termination codon at 2003
UT-1019 ²	5	1	0	2	nd	0.95	4132 del3	11	1302	deletion of thr ₁₃₀₂
UT-2027 ²	4	4	0	1	0.39	0.79	ND			
UT-2263 ²	3	2	0	1	nd	0.9	ND			
UT-2171 ²	5	4	2	0	nd	nd	ND			

¹Families screened for complete coding sequence and with informative cDNA sample.
²Families screened for all BRCA2 exons except 15 and for which there was no informative cDNA sample available.
IR — inferred regulatory mutation
ND — non detected
nd — not determined
FBC — Female Breast Cancer
Ov — Ovarian Cancer
MBC — Male Breast Cancer

Sequence alterations were identified in 9 of 18 kindreds. All except one involved nucleotide deletions that altered the reading frame, leading to truncation of the predicted BRCA2 protein. The single exception contained a deletion of three nucleotides (kindred 1019). All nine mutations differed from one another. A subset of kindreds was tested for transcript loss. cDNA samples were available for a group of nine kindreds, but three of the nine kindreds in the group contained frameshift mutations. Specific polymorphic sites known to be heterozygous in genomic DNA were examined in cDNA from kindred individuals. The appearance of hemizygosity at these polymorphic sites was interpreted as evidence for a mutation leading to reduction in mRNA levels. In only one of the six cases with no detectable sequence alteration (kindred 2367) could such a regulatory mutation be inferred. In addition, one of the three kindreds with a frameshift mutation (kindred 2044) displayed signs of transcript loss. This implies that some mutations in the BRCA2 coding sequence may destabilize the transcript in addition to disrupting the protein sequence. Such mutations have been observed in BRCA1 (Friedman et al., 1995). Thus, 56% of the kindreds (10 of 18) contained an altered BRCA2 gene.

Role of BRCA2 in Cancer. Most tumor suppressor genes identified to date give rise to protein products that are absent, nonfunctional, or reduced in function. The majority of TP53 mutations are missense; some of these have been shown to produce abnormal p53 molecules that interfere with the function of the wildtype product (Shaulian et al., 1992; Srivastava et al., 1993). A similar dominant negative mechanism of action has been proposed for some adenomatous polyposis coli (APC) alleles that produce truncated molecules (Su et al., 1993), and for point mutations in the Wilms' tumor gene (WT1) that alter DNA binding of the protein (Little et al., 1993). The nature of the mutations observed in the BRCA2 coding sequence is consistent with production of either dominant negative proteins or nonfunctional proteins.

EXAMPLE 5

Analysis of the BRCA2 Gene

The structure and function of BRCA2 gene are determined according to the following methods.

Biological Studies. Mammalian expression vectors containing BRCA2 cDNA are constructed and transfected into appropriate breast carcinoma cells with lesions in the gene. Wild-type BRCA2 cDNA as well as altered BRCA2 cDNA are utilized. The altered BRCA2 cDNA can be obtained from altered BRCA2 alleles or produced as described below. Phenotypic reversion in cultures (e.g., cell morphology, doubling time, anchorage-independent growth) and in animals (e.g., tumorigenicity) is examined. The studies will employ both wild-type and mutant forms (Section B) of the gene.

Molecular Genetics Studies. In vitro mutagenesis is performed to construct deletion mutants and missense mutants (by single base-pair substitutions in individual codons and cluster charged→alanine scanning mutagenesis). The mutants are used in biological, biochemical and biophysical studies.

Mechanism Studies. The ability of BRCA2 protein to bind to known and unknown DNA sequences is examined. Its ability to transactivate promoters is analyzed by transient reporter expression systems in mammalian cells. Conventional procedures such as particle-capture and yeast two-hybrid system are used to discover and identify any functional partners. The nature and functions of the partners are characterized. These partners in turn are targets for drug discovery.

Structural Studies. Recombinant proteins are produced in *E. coli*, yeast, insect and/or mammalian cells and are used in crystallographical and NMR studies. Molecular modeling of the proteins is also employed. These studies facilitate structure-driven drug design.

6,033,857

49

EXAMPLE 6

Two Step Assay to Detect the Presence of BRCA2 in a Sample

Patient sample is processed according to the method disclosed by Antonarakis et al. (1985), separated through a 1% agarose gel and transferred to nylon membrane for Southern blot analysis.

Membranes are UV cross linked at 150 mJ using a GS Gene Linker (Bio-Rad). A BRCA2 probe selected from the sequence shown in FIG. 3 is subcloned into pTZ18U. The phagemids are transformed into *E. Coli* MV 1190 infected with M13KO7 helper phage (Bio-Rad, Richmond, Calif.). Single stranded DNA is isolated according to standard procedures (see Sambrook et al., 1989).

Blots are prehybridized for 15–30 min at 65° C. in 7% sodium dodecyl sulfate (SDS) in 0.5 M NaPO₄. The methods follow those described by Nguyen et al., 1992. The blots are hybridized overnight at 65° C. in 7% SDS, 0.5 M NaPO₄ with 25–50 ng/ml single stranded probe DNA. Post-hybridization washes consist of two 30 min washes in 5% SDS, 40 mM NaPO₄ at 65° C., followed by two 30 min washes in 1% SDS, 40 mM NaPO₄ at 65° C.

Next the blots are rinsed with phosphate buffered saline (pH 6.8) for 5 min at room temperature and incubated with 0.2% casein in PBS for 30–60 min at room temperature and rinsed in PBS for 5 min. The blots are then preincubated for 5–10 minutes in a shaking water bath at 45° C. with hybridization buffer consisting of 6 M urea, 0.3 M NaCl, and 5×Denhardt's solution (see Sambrook, et al., 1989). The buffer is removed and replaced with 50–75 μl/cm² fresh hybridization buffer plus 2.5 nM of the covalently cross-linked oligonucleotide-alkaline phosphatase conjugate with the nucleotide sequence complementary to the universal primer site (UP-AP, Bio-Rad). The blots are hybridized for 20–30 min at 45° C. and post hybridization washes are incubated at 45° C. as two 10 min washes in 6 M urea, 1×standard saline citrate (SSC), 0.1% SDS and one 10 min wash in 1×SSC, 0.1% Triton®X-100. The blots are rinsed for 10 min at room temperature with 1×SSC.

Blots are incubated for 10 min at room temperature with shaking in the substrate buffer consisting of 0.1 M diethanolamine, 1 mM MgCl₂, 0.02% sodium azide, pH 10.0. Individual blots are placed in heat sealable bags with substrate buffer and 0.2 mM AMPPD (3-(2'-spiroadamantane)-4-methoxy-4-(3'-phosphoryloxy)phenyl-1,2-dioxetane, disodium salt, Bio-Rad).

After a 20 min incubation at room temperature with shaking, the excess AMPPD solution is removed. The blot is exposed to X-ray film overnight. Positive bands indicate the presence of BRCA2.

EXAMPLE 7

Generation of Polyclonal Antibody against BRCA2

Segments of BRCA2 coding sequence are expressed as fusion protein in *E. coli*. The overexpressed protein is purified by gel elution and used to immunize rabbits and mice using a procedure similar to the one described by Harlow and Lane, 1988. This procedure has been shown to generate Abs against various other proteins (for example, see Kraemer et al., 1993).

Briefly, a stretch of BRCA2 coding sequence selected from the sequence shown in FIG. 3 is cloned as a fusion protein in plasmid PET5A (Novagen, Inc., Madison, Wis.).

50

After induction with IPTG, the overexpression of a fusion protein with the expected molecular weight is verified by SDS/PAGE. Fusion protein is purified from the gel by electroelution. The identification of the protein as the BRCA2 fusion product is verified by protein sequencing at the N-terminus. Next, the purified protein is used as immunogen in rabbits. Rabbits are immunized with 100 μg of the protein in complete Freund's adjuvant and boosted twice in 3 week intervals, first with 100 μg of immunogen in incomplete Freund's adjuvant followed by 100 μg of immunogen in PBS. Antibody containing serum is collected two weeks thereafter.

This procedure is repeated to generate antibodies against the mutant forms of the BRCA2 gene. These antibodies, in conjunction with antibodies to wild type BRCA2, are used to detect the presence and the relative level of the mutant forms in various tissues and biological fluids.

EXAMPLE 8

Generation of Monoclonal Antibodies Specific for BRCA2

Monoclonal antibodies are generated according to the following protocol. Mice are immunized with immunogen comprising intact BRCA2 or BRCA2 peptides (wild type or mutant) conjugated to keyhole limpet hemocyanin using glutaraldehyde or EDC as is well known.

The immunogen is mixed with an adjuvant. Each mouse receives four injections of 10 to 100 μg of immunogen and after the fourth injection blood samples are taken from the mice to determine if the serum contains antibody to the immunogen. Serum titer is determined by ELISA or RIA. Mice with sera indicating the presence of antibody to the immunogen are selected for hybridoma production.

Spleens are removed from immune mice and a single cell suspension is prepared (see Harlow and Lane, 1988). Cell fusions are performed essentially as described by Kohler and Milstein, 1975. Briefly, P3.65.3 myeloma cells (American Type Culture Collection, Rockville, Md.) are fused with immune spleen cells using polyethylene glycol as described by Harlow and Lane, 1988.

Cells are plated at a density of 2×10⁵ cells/well in 96 well tissue culture plates. Individual wells are examined for growth and the supernatants of wells with growth are tested for the presence of BRCA2 specific antibodies by ELISA or RIA using wild type or mutant BRCA2 target protein. Cells in positive wells are expanded and subcloned to establish and confirm monoclonality.

Clones with the desired specificities are expanded and grown as ascites in mice or in a hollow fiber system to produce sufficient quantities of antibody for characterization and assay development.

EXAMPLE 9

Sandwich Assay for BRCA2

Monoclonal antibody is attached to a solid surface such as a plate, tube, bead, or particle.

Preferably, the antibody is attached to the well surface of a 96-well ELISA plate. 100 μl sample (e.g., serum, urine, tissue cytosol) containing the BRCA2 peptide/protein (wild-type or mutant) is added to the solid phase antibody. The sample is incubated for 2 hrs at room temperature. Next the sample fluid is decanted, and the solid phase is washed with buffer to remove unbound material. 100 μl of a second

6,033,857

51

monoclonal antibody (to a different determinant on the BRCA2 peptide/protein) is added to the solid phase. This antibody is labeled with a detector molecule (e.g., ¹²⁵I, enzyme, fluorophore, or a chromophore) and the solid phase with the second antibody is incubated for two hrs at room temperature. The second antibody is decanted and the solid phase is washed with buffer to remove unbound material.

The amount of bound label, which is proportional to the amount of BRCA2 peptide/protein present in the sample, is quantitated. Separate assays are performed using monoclonal antibodies which are specific for the wild-type BRCA2 as well as monoclonal antibodies specific for each of the mutations identified in BRCA2.

EXAMPLE 10

The 6174delT Mutation is Common in Ashkenazi Jewish Women Affected by Breast Cancer

The 6174delT mutation (see Table 3) has been found to be present in many cases of Ashkenazi Jewish women who have had breast cancer (Neuhausen et al., 1996). Two groups of probands comprised the ascertainment for this study. The first group was ascertained based on both age-of-onset and a positive family history. The first group consisted of probands affected with breast cancer on or before 41 years of age with or without a family history of breast cancer. Inclusion criteria for the second group were that the proband was affected with breast cancer between the ages of 41 and 51 with one or more first degree relatives affected with breast or ovarian cancer on or before the age of 50; or the proband was affected with breast cancer between the ages of 41 and 51 with two or more second degree relatives affected with breast or ovarian cancer, 1 on or before age 50; or the proband was affected between the ages of 41 and 51 with both primary breast and primary ovarian cancer. Probands were ascertained through medical oncology and genetic counseling clinics, with an effort to offer study participation to all eligible patients. Family history was obtained by a self-report questionnaire. Histologic confirmation of diagnosis was obtained for probands in all cases. Religious background was confirmed on all probands by self report or interview.

Mutation Detection

The BRCA2 6174delT mutation was detected by amplifying genomic DNA from each patient according to standard polymerase chain reaction (PCR) procedures (Saiki et al., 1985; Mullis et al., 1986; Weber and May, 1989). The primers used for the PCR are:

BC11-RP: GGGAAGCTTCATAAGTCAGTC (SEQ ID NO:115)

(forward primer) and

BC11-LP: TTGTGAATGAAGCATCTGATACC (SEQ ID NO:116)

(reverse primer).

The reactions were performed in a total volume of 10.0 µl containing 20 µg DNA with annealing at 55° C. This produces a PCR product 97 bp long in wild-type samples and 96 bp long when the 6174delT mutation is present. The radiolabeled PCR products were electrophoresed on standard 6% polyacrylamide denaturing sequencing gels at 65 W for 2 hours. The gels were then dried and autoradiographed. All the cases exhibiting the 1 bp deletion were sequenced to confirm the 6174delT mutation. For sequencing, half of the samples were amplified with one set of PCR primers and the coding strand was sequenced and the other half of the samples were amplified with a second set of PCR primers

52

and the noncoding strand was sequenced. For one set the PCR primers were:

TD-SFB: AATGATGAATGTAGCACGC (SEQ ID NO:117)

(forward primer) and

CGORF-RH: GTCTGAATGTTCTGTACT (SEQ ID NO:118)

(reverse primer).

This results in an amplified product of 342 bp in wild-type and 341 bp for samples containing the 6174delT mutation. For this set of samples the amplified DNA was sequenced using the CGORF-RH primer for the sequencing primer. The other half of the samples were amplified using the BC11 -RP forward primer and the CGORF-RH reverse primer resulting in a fragment of 183 bp in wild-type samples and 182 bp in samples containing the 6174delT mutation. This was sequenced using BC11-RP as the sequencing primer.

Results

Six out of eighty women of Ashkenazi Jewish ancestry with breast cancer before the age of 42 had the 6174delT mutation. This compares to zero cases of the mutation being present in a control group of non-Jewish women who had breast cancer before the age of 42. These cases were ascertained without regard to family history. Table 4 shows the results of the study. Four of the six cases with the 6174delT mutation had a family history of breast or ovarian cancer in a first or second degree relative. In each of two kindreds where multiple samples were available for analysis, the 6174delT mutation co-segregated with two or more cases of breast or ovarian cancer. A second cohort of 27 Ashkenazim with breast cancer at age 42–50 and a history of at least one additional relative affected with breast or ovarian cancer provided an additional estimate of the frequency of the 6174delT mutation. In this group of 27 women, two were heterozygous for the BRCA2 6174delT mutation. One of these individuals had first degree relatives with both ovarian and breast cancer. From the data presented, and assuming a penetrance similar to BRCA1 mutations (Offit et al., 1996; Langston et al., 1996), the frequency of the 6174delT mutation in Ashkenazim can be estimated to be approximately 3 per thousand. However, if the penetrance of this mutation is lower than BRCA1, then the frequency of this mutation will be higher. A more precise estimate of the carrier frequency of the 6174delT mutation in individuals of Ashkenazi Jewish ancestry will emerge from large-scale population studies.

TABLE 4

Group	Number of subjects tested, n =	Number with 6174delT, n =	%
Group 1a			
Diagnosis before age 42, Non-Jewish ^a	93	0	(0)
Group 1b			
Diagnosis before age 42, Jewish ^a	80	6	(8)
Before age 37	40	4	(10)
age 37–41	40	2	(5)

TABLE 4-continued

Group	Number of subjects tested, n =	Number with 6174delT, n =	%
Group 2			
Diagnosis ages 42–50 and family history positive ^b	27	2	(27)

Key:
^aAscertained regardless of family history
^bFamily history for this group was defined as one first degree or two second degree relatives diagnosed with breast or ovarian cancer, one before age 50.

EXAMPLE 11

BRCA2 Shows a Low Somatic Mutation Rate in Breast Carcinoma and Other Cancers Including Ovarian and Pancreatic Cancers

BRCA2 is a tumor suppressor gene. A homozygous deletion of this gene may lead to breast cancer as well as other cancers. A homozygous deletion in a pancreatic xenograft was instrumental in the effort to isolate BRCA2 by positional cloning. Cancer may also result if there is a loss of one BRCA2 allele and a mutation in the remaining allele (loss of heterozygosity or LOH).

Mutations in both alleles may also lead to development of cancer. For studies here, an analysis of 150 cell lines derived from different cancers revealed no cases in which there was a homozygous loss of the BRCA2 gene. Because homozygous loss is apparently rare, investigations were made to study smaller lesions such as point mutations in BRCA2. Since compound mutant heterozygotes and mutant homozygotes are rare, tumor suppressor gene inactivation nearly always involves LOH. The remaining allele, if inactive, typically contains disruptive mutations. To identify these it is useful to preselect tumors or cell lines that exhibit LOH at the locus of interest.

Identification of tumors and cell lines that exhibit LOH

A group of 104 primary breast tumor samples and a set of 269 cell lines was tested for LOH in the BRCA2 region. For primary tumors, amplifications of three short tandem repeat markers (STRs) were compared quantitatively using fluorescence. Approximately 10 ng of genomic DNA was amplified by PCR with the following three sets of fluorescently tagged STRs:

- (1) mM4247.4A.2F1 ACCATCAAACACATCATCC (SEQ ID NO: 119)
mM4247.4A.2R2 AGAAAGTAACTGGAGGGAG (SEQ ID NO: 120)
- (2) STR257-FC CTCTGAAACTGTTCCCTTGG (SEQ ID NO: 121)
STR257-RD TAATGGTGCTGGGATATTGG (SEQ ID NO: 122)
- (3) mMB561A-3.1FA2 GAATGTCGAAGAGCTTGTC (SEQ ID NO: 123)
mMB561A-3.1RB AAACATACGCTTAGCCAGAC (SEQ ID NO: 124)

The PCR products were resolved using an ABI 377 sequencer and quantified with Genescan software (ABI). For tumors, clear peak height differences between alleles amplified from normal and tumor samples were scored as having LOH. For cell lines, if one STR was heterozygous, the sample was scored as non-LOH. In only one case was a cell line or tumor miscalled based on later analysis of single base polymorphisms. The heterozygosity indices for the markers

are: STR4247 =0.89; STR257=0.72; STR561A=0.88 (S. Neuhausen, personal communication; B. Swedlund, unpublished data). Based on their combined heterozygosity indices, the chance that the markers are all homozygous in a particular individual (assuming linkage equilibrium) is only one in 250. Due to the presence of normal cells in the primary tumor sample, LOH seldom eliminates the signal entirely from the allele lost in the tumor. Rather, the relative intensities of the two alleles are altered. This can be seen clearly by comparing the allelic peak heights from normal tissue with peak heights from the tumor (FIGS. 5A–5D). Based on this analysis, 30 tumors (29%) were classified as having LOH at the BRCA2 locus (Table 5), a figure that is similar to previous estimates (Collins et al., 1995; Cleton-Jansen et al., 1995).

LOH was assessed in the set of cell lines in a different fashion. Since homozygosity of all three STRs was improbable, and since normal cells were not present, apparent homozygosity at all STRs was interpreted as LOH in the BRCA2 region. Using this criterion, 85/269 of the cell lines exhibited LOH (see Table 5). The frequencies varied according to the particular tumor cell type under consideration. For example, 4% ovarian cell lines and 31/62 lung cancer lines displayed LOH compared with 17/81 melanoma lines and 2/11 breast cancer lines.

Sequence Analysis of LOH Primary Breast Tumors and Cell Lines

The 30 primary breast cancers identified above which showed LOH in the BRCA2 region were screened by DNA sequence analysis for sequence variants. Greater than 95% of the coding sequence and splice junctions was examined. DNA sequencing was carried out either on the ABI 377 (Applied Biosystems Division, Perkin-Elmer) or manually. For the radioactive mutation screen, the amplified products were purified by Qiagen beads (Qiagen, Inc.). DNA sequence was generated using the Cyclist sequencing kit (Stratagene) and resolved on 6% polyacrylamide gels. In parallel, non-radioactive sequencing using fluorescent labeling dyes was performed using the TaqFS sequencing kit followed by electrophoresis on ABI 377 sequencers. Samples were gridded into 96-well trays to facilitate PCR and sequencing. Dropouts of particular PCR and sequencing reactions were repeated until >95% coverage was obtained for every sample. Sequence information was analyzed with the Sequencher software (Gene Codes Corporation). All detected mutations were confirmed by sequencing a newly amplified PCR product to exclude the possibility that the sequence alteration was due to a PCR artifact.

TABLE 5

Type	# LOH/# Screened	Percentage LOH	# Sequenced
Astrocytoma	6/19	32%	6
Bladder	6/17	35%	4
Breast	2/11	18%	2
Colon	2/8	25%	2
Glioma	11/36	31%	5
Lung	31/62	50%	20
Lymphoma	0/4	0%	0
Melanoma	17/81	21%	9
Neuroblastoma	1/10	10%	1
Ovarian	4/6	67%	4
Pancreatic	1/3	33%	1
Prostate	0/2	0%	0
Renal	4/10	40%	4
Total	85/269	33% (avg. = 28%)	58
Primary Breast	30/104	29%	42

LOH analysis of cell lines and primary breast tumors. Percentage LOH was calculated two ways: as total and as a mean of percentages (avg.).

Of the 30 samples, two specimens contained frameshift mutations, one a nonsense mutation, and two contained missense changes (although one of these tumors also contained a frameshift). The nonsense mutation would delete 156 codons at the C-terminus suggesting that the C-terminal end of BRCA2 is important for tumor suppressor activity. All sequence variants were also present in the corresponding normal DNA from these cancer patients. To exclude the unlikely possibility that preselection for LOH introduced a systematic bias against detecting mutations (e.g., dominant behavior of mutations, compound heterozygotes), 12 samples shown to be heterozygous at BRCA2 were also screened. Three of these revealed missense changes that were also found in the normal samples. Thus, in a set of 42 breast carcinoma samples, 30 of which displayed LOH at the BRCA2 locus, no somatic mutations were identified. The frameshift and nonsense changes are likely to be predisposing mutations that influenced development of breast cancer in these patients. The missense variants are rare; they were each observed only once during analysis of 115 chromosomes. From these data it is not possible to distinguish between rare neutral polymorphisms and predisposing mutations.

Of the 85 cell lines which displayed LOH (see Table 5), 58 were also screened for sequence changes. Greater than 95% of the coding sequence of each sample was screened. Only a single frameshift mutation was identified by this DNA sequence analysis. This mutation (6174delT) was present in a pancreatic cancer line and it is identical to one found in the BT111 primary tumor sample and to a previously detected germline frameshift (Tavtigian et al., 1996). This suggests that this particular frameshift may be a relatively common germline BRCA2 mutation. In addition, a number of missense sequence variants were detected (Tables 6A and 6B).

Detection of a probable germline BRCA2 mutation in a pancreatic tumor cell line suggests that BRCA2 mutations may predispose to pancreatic cancer, a possibility that has not been explored thoroughly. This mutation also adds weight to the involvement of BRCA2 in sporadic pancreatic cancer, implied previously by the homozygous deletion observed in a pancreatic xenograft (Schutte et al., 1995). Because only three pancreatic cell lines were examined in our study, further investigation of BRCA2 mutations in pancreatic cancers is warranted.

TABLE 6A

Sample	Type	LOH	Change	Effect	Germline
4H5	Renal	yes	G451C	Ala→Pro	
4G1	Ovarian	yes	A1093C	Asn→His	
2F8	Lung	yes	G1291C	Val→Leu	
BT110	Primary breast	yes	1493delA	Frameshift	yes
4F8	Ovarian	yes	C2117T	Thr→Ile	
BT163	Primary breast	no	A2411C	Asp→Ala	yes
1D6	Bladder	no	G4813A	Gly→Arg	
BT333	Primary breast	no	T5868G	Asn→Lys	yes
2A2	Glioma	yes	C5972T	Thr→Met	
2I4	Lung	yes	C5972T	Thr→Met	
BT111	Primary breast	yes	6174delT	Frameshift	yes
4G3	Pancreatic	yes	6174delT	Frameshift	
1B7	Astrocytoma	yes	C6328T	Arg→Cys	
BT118	Primary breast	no	G7049T	Gly→Val	yes
BT115	Primary breast	yes	G7491C	Gln→His	yes
3D5	Melanoma	yes	A9537G	Ile→Met	

TABLE 6A-continued

Sample	Type	LOH	Change	Effect	Germline
BT85	Primary breast	yes	A10204T	Lys→Stop	yes
1E4	Breast	yes	C10298G	Thr→Arg	
BT110	Primary breast	yes	A10462G	Ile→Val	yes

Germline mutations identified in BRCA2. Listed are the mutation positions based on the Genbank entry of BRCA2 (Schutte et al., 1995).

TABLE 6B

Position	Change	Effect	Frequency
5'UTR(203)	G/A	—	0.32 (0.26)
PM(1342)	C/A	His→Asn	0.32 (0.37)
PM(2457)	T/C	silent	0.04 (0.05)
PM(3199)	A/G	Asn→Asp	0.04 (0.08)
PM(3624)	A/G	silent	0.35
PM(3668)	A/G	Asn→Ser	0 (0.15)
PM(4035)	T/C	silent	0.24 (0.10)
PM(7470)	A/G	silent	0.26 (0.15)
1593	A→G	silent	<0.01
4296	G→A	silent	<0.01
5691	A→G	silent	<0.01
6051	A→G	silent	<0.01
6828	T→C	silent	<0.01
6921	T→C	silent	<0.01

Common polymorphisms and silent substitutions detected in BRCA2 by DNA sequencing. Since some rare silent variants may affect gene function (e.g., splicing (Richard and Beckmann, 1995)), these are not preceded by "PM". The frequencies of polymorphisms shown involve the second of the nucleotide pair. Frequencies reported in a previous study are shown in parentheses (Tavtigian et al., 1996). Numbering is as in Table 6A.

Industrial Utility
As previously described above, the present invention provides materials and methods for use in testing BRCA2 alleles of an individual and an interpretation of the normal or predisposing nature of the alleles. Individuals at higher than normal risk might modify their lifestyles appropriately. In the case of BRCA2, the most significant non-genetic risk factor is the protective effect of an early, full term pregnancy. Therefore, women at risk could consider early childbearing or a therapy designed to simulate the hormonal effects of an early full-term pregnancy. Women at high risk would also strive for early detection and would be more highly motivated to learn and practice breast self examination. Such women would also be highly motivated to have regular mammograms, perhaps starting at an earlier age than the general population. Ovarian screening could also be undertaken at greater frequency. Diagnostic methods based on sequence analysis of the BRCA2 locus could also be applied to tumor detection and classification. Sequence analysis could be used to diagnose precursor lesions. With the evolution of the method and the accumulation of information about BRCA2 and other causative loci, it could become possible to separate cancers into benign and malignant.

Women with breast cancers may follow different surgical procedures if they are predisposed, and therefore likely to have additional cancers, than if they are not predisposed. Other therapies may be developed, using either peptides or small molecules (rational drug design). Peptides could be the missing gene product itself or a portion of the missing gene product. Alternatively, the therapeutic agent could be another molecule that mimics the deleterious gene's

function, either a peptide or a nonpeptidic molecule that seeks to counteract the deleterious effect of the inherited locus. The therapy could also be gene based, through introduction of a normal BRCA2 allele into individuals to make a protein which will counteract the effect of the deleterious allele. These gene therapies may take many forms and may be directed either toward preventing the tumor from forming, curing a cancer once it has occurred, or stopping a cancer from metastasizing.

It will be appreciated that the methods and compositions of the instant invention can be incorporated in the form of a variety of embodiments, only a few of which are disclosed herein. It will be apparent to the artisan that other embodiments exist and do not depart from the spirit of the invention. Thus, the described embodiments are illustrative and should not be construed as restrictive.

LIST OF REFERENCES

American Cancer Society, Cancer Facts & Figures—1992. (American Cancer Society, Atlanta, Ga).
Anand, R. (1992). *Techniques for the Analysis of Complex Genomes*, (Academic Press).
Anderson, et al. (1980). *Proc. Natl. Acad. Sci. USA* 77:5399–5403.
Anderson, D. E. (1972). *J. Natl. Cancer Inst.* 48:1029–1034.
Anderson, J. A., et al. (1992). *J. Otolaryngology* 21:321.
Antonarakis, S. E., et al. (1985). *New Eng. J. Med.* 313:842–848.
Ausubel, F. M., et al. (1992). *Current Protocols in Molecular Biology*, (J. Wiley and Sons, N.Y.)
Beaucage & Carruthers (1981). *Tetra. Letts.* 22:1859–1862.
Berkner (1992). *Curr. Top. Microbiol. Immunol.* 158:39–61.
Berkner, et al. (1988). *BioTechniques* 6:616–629.
Bishop, D. T., et al. (1988). *Genet. Epidemiol.* 5:151–169.
Bishop, D. T. and Gardner, E. J. (1980). In: *Banbury Report 4: Cancer Incidence in Defined Populations* (J. Cairns, J. L. Lyon, M. Skolnick, eds.), Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 309–408.
Botstein, et al. (1980). *Am. J. Hum. Genet.* 32:314–331.
Brandyopadhyay and Temin (1984). *Mol. Cell. Biol.* 4:749–754.
Breakfield and Geller (1987). *Mol. Neurobiol.* 1:337–371.
Brinster, et al. (1981). *Cell* 27:223–231.
Buchsachacher and Panganiban (1992). *J. Virol.* 66:2731–2739.
Buckler, et al. (1991). *Proc. Natl. Acad. Sci. USA* 88:4005–4009.
Cannon-Albright, L., et al. (1994). *Cancer Research* 54:2378–2385.
Capecci, M. R. (1989). *Science* 244:1288.
Cariello (1988). *Human Genetics* 42:726.
Church, D. M., et al., (1993). *Hum. Molec. Genet.* 2:1915.
Claus, E., et al. (1991). *Am. J. Hum. Genet.* 48:232–242.
Cleton-Jansen, A. M., et al. (1995). *Br. J. Cancer* 72:1241–1244.
Collins, N., et al. (1995). *Oncogene* 10:1673–1675.
Conner, B. J., et al. (1983). *Proc. Natl. Acad. Sci. USA* 80:278–282.
Constantini and Lacy (1981). *Nature* 294:92–94.
Cotten, et al. (1990). *Proc. Natl. Acad. Sci. USA* 87:4033–4037.
Cotton, et al. (1988). *Proc. Natl. Acad. Sci. USA* 85:4397.
Culver, et al. (1992). *Science* 256:1550–1552.
Curiel, et al. (1991a). *Proc. Natl. Acad. Sci. USA* 88:8850–8854.
Curiel, et al. (1991 b). *Hum. Gene Ther.* 3:147–154.
Deutscher, M. (1990). *Meth. Enzymology* 182 (Academic Press, San Diego, Calif.).

Donehower, L. A., et al. (1992). *Nature* 356:215.
Enhancers and Eukaryotic Gene Expression, Cold Spring Harbor Press, Cold Spring Harbor, N.Y. (1983).
Easton et al. (1993). *Am. J. Hum. Gen.* 52:678–701.
Erickson, J. et al., (1990). *Science* 249:527–533.
Felgner, et al. (1987). *Proc. Natl. Acad. Sci. USA* 84:7413–7417.
Fiers, et al. (1978). *Nature* 273:113.
Fink, et al. (1992). *Hum. Gene Ther.* 3:11–19.
Finkelstein, J., et al. (1990). *Genomics* 7:167–172.
Freese, et al. (1990). *Biochem. Pharmacol.* 40:2189–2199.
Friedman, T. (1991). In *Therapy for Genetic Diseases*, T. Friedman, ed., Oxford University Press, pp. 105–121.
Friedman, L. S., et al. (1995). *Am. J. Hum. Genet.* 57:1284–1297.
Futreal (1993). Ph.D. Thesis, University of North Carolina, Chapel Hill.
Futreal et al. (1994). *Science* 266:120–122.
Glover, D. (1985). *DNA Cloning*, I and II (Oxford Press).
Go, R. C. P., et al. (1983). *J. Natl. Cancer Inst.* 71:455–461.
Goding (1986). *Monoclonal Antibodies: Principles and Practice*, 2d ed. (Academic Press, N.Y.).
Godowski, et al. (1988). *Science* 241:812–816.
Gordon, et al. (1980). *Proc. Natl. Acad. Sci. USA* 77:7380–7384.
Gorziglia and Kapikian (1992). *J. Virol.* 66:4407–4412.
Graham and van der Eb (1973). *Virology* 52:456–467.
Grompe, M., (1993). *Nature Genetics* 5:111–117.
Grompe, M., et al., (1989). *Proc. Natl. Acad. Sci. USA* 86:5855–5892.
Guthrie, G. & Fink, G. R. (1991). *Guide to Yeast Genetics and Molecular Biology* (Academic Press).
Hall, J. M., et al. (1990). *Science* 250:1684–1689.
Harlow & Lane (1988). *Antibodies: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.).
Hasty, P., K., et al. (1991). *Nature* 350:243.
Helseth, et al. (1990). *J. Virol.* 64:2416–2420.
Hodgson, J. (1991). *BioTechnology* 9:19–21.
Huse, et al. (1989). *Science* 246:1275–1281.
Innis et al. (1990). *PCR Protocols. A Guide to Methods and Applications* (Academic Press, San Diego, Calif.).
Jablonski, E., et al. (1986). *Nuc. Acids Res.* 14:6115–6128.
Jakoby, W. B. and Pastan, I. H. (eds.) (1979). *Cell Culture. Methods in Enzymology*, volume 58 (Academic Press, Inc., Harcourt Brace Jovanovich (N.Y.)).
Jeffreys, et al. (1985). *Nature* 314:67–73.
Johnson, et al. (1992). *J. Virol.* 66:2952–2965.
Kamb, A. et al. (1994a). *Science* 264:436–440.
Kamb, A. et al. (1994b). *Nature Genetics* 8:22.
Kandpal, et al. (1990). *Nucl. Acids Res.* 18:1789–1795.
Kaneda, et al. (1989). *J. Biol. Chem.* 264:12126–12129.
Kanehisa (1984). *Nucl. Acids Res.* 12:203–213.
Kinszler, K. W., et al. (1991). *Science* 251:1366–1370.
Knudson, A. G. (1993). *Nature Genet.* 5:103.
Kohler, G. and Milstein, C. (1975). *Nature* 256:495–497.
Kraemer, F. B. et al. (1993). *J. Lipid Res.* 34:663–672.
Kubo, T., et al. (1988). *FEBS Letts.* 241:119.
Landegren, et al. (1988). *Science* 242:229.
Langston, A. A., et al. (1996). *N. Engl. J. Med.* 334:137–142.
Lim, et al. (1992). *Circulation* 83:2007–2011.
Lindsay, S., et al. (1987). *Nature* 327:336–368.
Litt, et al. (1989). *Am. J. Hum. Genet.* 44:397–401.
Little, M. H., et al. (1993). *Hum. Mol. Genet.* 2:259.
Lovett, et al. (1991). *Proc. Natl. Acad. Sci. USA* 88:9628–9632.
Madzak, et al. (1992). *J. Gen. Virol.* 73:1533–1536.

6,033,857

59

Malkin, D., et al. (1990). *Science* 250:1233-1238.
Maniatis, T. et al. (1982). *Molecular Cloning. A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.).
Mann and Baltimore (1985). *J. Virol.* 54:401-407.
Margaritte, et al. (1992). *Am. J. Hum. Genet.* 50:1231-1234.
Margolske (1992). *Curr. Top. Microbiol. Immunol.* 158:67-90.
Martin, R., et al. (1990). *Bio Techniques* 9:762-768.
Matteucci, M. D. and Caruthers, M. H. (1981). *J. Am. Chem. Soc.* 103:3185.
Matthews & Kricka (1988). *Anal. Biochem.* 169:1.
Merrifield (1963). *J. Am. Chem. Soc.* 85:2149-2156.
Mettlin, C., et al. (1990). *American Journal of Epidemiology* 131:973-983.
Metzger, et al. (1988). *Nature* 334:31-36.
Miller (1992). *Curr. Top. Microbiol. Immunol.* 158:1-24.
Miller, et al. (1985). *Mol. Cell. Biol.* 5:431-437.
Miller, et al. (1988). *J. Virol.* 62:4337-4345.
Mittlin (1989). *Clinical Chem.* 35:1819.
Modrich, P. (1991). *Ann. Rev. Genet.* 25:229-253.
Mombaerts, P., et al. (1992). *Cell* 68:869.
Moss (1992). *Curr. Top. Microbiol. Immunol.* 158:25-38.
Mullis, K., et al. (1986). *Cold Spring Harbor Symp. Quant. Biol.* 51:263-273.
Muzyczka (1992). *Curr. Top. Microbiol. Immunol.* 158:97-123.
Nabel (1992). *Hum. Gene Ther.* 3:399-410.
Nabel, et al. (1990). *Science* 249:1285-1288.
Nakamura, B., et al. (1987). *Science* 235:1616-1622.
Narod, S. A., et al. (1991). *The Lancet* 338:82-83.
Neuhausen, S., et al. (1996). *Nature Genetics* 13:(in press, May 1996 issue).
Newman, B., et al. (1988). *Proc. Natl. Acad. Sci. USA* 85:3044-3048.
Newton, C. R., Graham, A., Heptinstall, L. E., Powell, S. J., Summers, C., Kalsheker, N., Smith, J. C., and Markham, A. F. (1989). *Nucl. Acids Res.* 17:2503-2516.
Nguyen, Q., et al. (1992). *BioTechniques* 13:116-123.
Novack, et al. (1986). *Proc. Natl. Acad. Sci. USA* 83:586.
Offit, K., et al. (1996). *Lancet* (In press).
Ohi, et al. (1990). *Gene* 89:279-282.
Orita, et al. (1989). *Proc. Natl. Acad. Sci. USA* 86:2776-2770.
Ott, J. (1985). *Analysis of Human Genetic Linkage, Johns Hopkins University Press, Baltimore, Md., pp. 1-216.*
Page, et al. (1990). *J. Virol.* 64:5370-5276.
Parimoo, S., et al. (1991). *Proc. Natl. Acad. Sci. USA* 88:9623-9627.
Pellicer, et al. (1980). *Science* 209:1414-1422.
Petrooulos, et al. (1992). *J. Virol.* 66:3391-3397.
Philpott, K. L., et al. (1992). *Science* 256:1448.
Pierce, et al. (1992). *Proc. Natl. Acad. Sci. USA* 89:2056-2060.
Quantin, et al. (1992). *Proc. Natl. Acad. Sci. USA* 89:2581-2584.
Rano & Kidd (1989). *Nucl. Acids Res.* 17:8392.
Richard, L. and Beckmann, J. S. (1995). *Nature Genetics* 10:259.
Rigby, P. W. J., et al. (1977). *J. Mol. Biol.* 113:237-251.
Rommens, J. M. et al. (1994). In *Identification of Transcribed Sequences*, (U. Hochgeschwender & K. Gardiner, Eds.), Plenum Press, New York, 65-79.
Rosenfeld, et al. (1992). *Cell* 68:143-155.
Saiki, R. K., et al. (1985). *Science* 230:1350-1354.
Sambrook, J., et al. (1989). *Molecular Cloning: A Laboratory Manual*, 2nd Ed. (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.).

60

Scharf (1986). *Science* 233:1076.
Schutte, M., et al. (1995). *Proc. Natl. Acad. Sci.* 92:5950-5954.
Scopes, R. (1982). *Protein Purification. Principles and Practice*, (Springer-Verlag, N.Y.).
Shaulian, E., et al. (1992). *Mol. Cell Biol.* 12:5581-92.
Sheffield, V. C., et al. (1989). *Proc. Natl. Acad. Sci. USA* 86:232-236.
Sheffield, V. C., et al. (1991). *Am. J. Hum. Genet.* 49:699-706.
Shenk, et al. (1975). *Proc. Natl. Acad. Sci. USA* 72:989.
Shimada, et al. (1991). *J. Clin. Invest.* 88:1043-1047.
Shinkai, Y., et al. (1992). *Cell* 68:855.
Shizuya, H., et al. (1992). *Proc. Natl. Acad. Sci. USA* 89:8794-8797.
Skolnick, M. H. and Wallace, B. R. (1988). *Genomics* 2:273-279.
Skolnick, M. H., et al. (1990). *Science* 250:1715-1720.
Smith, S. A., et al. (1992). *Nature Genetics* 2:128-131.
Snouwraert, J. N., et al. (1992). *Science* 257:1083.
Sorge, et al. (1984). *Mol. Cell. Biol.* 4:1730-1737.
Srivastava, S., et al. (1993). *Cancer Res.* 53:4452-5.
Sternberg (1990). *Proc. Natl. Acad. Sci. USA* 87:103-107.
Sternberg, et al. (1990). *The New Biologist* 2:151-162.
Stewart, et al. (1992). *Hum. Gene Ther.* 3:267-275.
Stone, S., et al. (1995). *Cancer Research* 55:2988-2994.
Stratford-Perricaudet, et al. (1990). *Hum. Gene Ther.* 1:241-256.
Swift, M., et al. (1991). *N. Engl. J. Med.* 325:1831-1836.
Swift, M., et al. (1976). *Cancer Res.* 36:209-215.
Su, L. K., et al. (1993). *Cancer Res.* 53:2728-31.
Tavtigian, S. V., et al. (1996). *Nature Genetics* 12:1-6.
Thomas, A. and Skolnick, M. H. (1994). *IMA Journal of Mathematics Applied in Medicine and Biology* (in press).
Tonolio, D., et al. (1990). Cold Spring Harbor Conference.
Valancius, V. & Smithies, O. (1991). *Mol. Cell Biol.* 11:1402.
Wagner, et al. (1990). *Proc. Natl. Acad. Sci. USA* 87:3410-3414.
Wagner, et al. (1991). *Proc. Natl. Acad. Sci. USA* 88:4255-4259.
Wang and Huang (1989). *Biochemistry* 28:9508-9514.
Wartell, R. M., et al. (1990). *Nucl. Acids Res.* 18:2699-2705.
Weber, J. L. (1990). *Genomics* 7:524-530.
Weber and May (1989). *Am. J. Hum. Genet.* 44:388-396.
Weber, J. L., et al. (1990). *Nucleic Acid Res.* 18:4640.
Wells, J. A. (1991). *Methods in Enzymol.* 202:390-411.
Wetmur & Davidson (1968). *J. Mol. Biol.* 31:349-370.
White, M. B., et al., (1992). *Genomics* 12:301-306.
White and Lalouel (1988). *Ann. Rev. Genet.* 22:259-279.
Wilkinson, et al. (1992). *Nucleic Acids Res.* 20:2233-2239.
Williams and Anderson (1984). *Genet. Epidemiol.* 1:7-20.
Wolff, et al. (1990). *Science* 247:1465-1468.
Wolff, et al. (1991). *BioTechniques* 11:474-485.
Wooster, R., et al. (1994). *Science* 265:2088.
Wu, et al. (1989a). *Genomics* 4:560-569.
Wu, et al. (1989b). *J. Biol. Chem.* 264:16985-16987.
Wu, et al. (1991). *J. Biol. Chem.* 266:14338-14342.
Zenke, et al. (1990). *Proc. Natl. Acad. Sci. USA* 87:3655-3659.
List of Patents and Patent Applications:
U.S. Pat. No. 3,817,837
U.S. Pat. No. 3,850,752
U.S. Pat. No. 3,939,350
U.S. Pat. No. 3,996,345
U.S. Pat. No. 4,275,149
U.S. Pat. No. 4,277,437

6,033,857

61

62

U.S. Pat. No. 4,366,241
U.S. Pat. No. 4,376,110
U.S. Pat. No. 4,486,530
U.S. Pat. No. 4,683,195
U.S. Pat. No. 4,683,202
U.S. Pat. No. 4,816,567
U.S. Pat. No. 4,868,105

U.S. Pat. No. 5,252,479
EPO Publication No. 225,807
European Patent Application Publication No. 0332435
Geysen, H., PCT published application WO 84/03564, pub-
5 lished Sep. 13, 1984
Hitzeman et al., EP 73,675A
PCT published application WO 93/07282

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(iii) NUMBER OF SEQUENCES: 124

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 11385 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

```
(ix) FEATURE:
      (A) NAME/KEY: CDS
      (B) LOCATION: 229..10482
```

(xi) SEQUENCE DESCRIPTION: SEO ID NO:1:

GGTGGCGCGA	GCTTCTGAAA	CTAGGCGGCA	GAGGCGGAGC	CGCTGTGGCA	CTGCTGCGCC	60
TCTGCTGCGC	CTCGGGTGTC	TTTTCGCGCG	GTGGGTCGCC	GCCGGGAGAA	GCGTGAGGGG	120
ACAGATTGTG	GACCGGCGCG	GTTTTTGTC	GCTTACTCCG	GCCAAAAAAG	AACTGCACCT	180
CTGAGACGGA	CTTATTACC	AAGCATTGGA	GGAATATCGT	AGGTAAAA	ATG CCT ATT	237
					Met Pro Ile	
					1	
GGA TCC AAA	GAG AGG CCA	ACA TTT TTT	GAA ATT TTT	AAG ACA CGC	TGC	285
Gly Ser Lys	Glu Arg Pro	Thr Phe Phe	Glu Ile Phe	Lys Thr Arg	Cys	
5		10		15		
AAC AAA GCA	GAT TTA GGA	CCA ATA AGT	CTT AAT TGG	TTT GAA GAA	CTT	333
Asn Lys Ala	Asp Leu Gly	Pro Ile Ser	Leu Asn Trp	Phe Glu Glu	Leu	
20		25		30	35	
TCT TCA GAA	GCT CCA CCC	TAT AAT TCT	GAA CCT GCA	GAA GAA TCT	GAA	381
Ser Ser Glu	Ala Pro Pro	Tyr Asn Ser	Glu Pro Ala	Glu Glu Ser	Glu	
	40		45	50		
CAT AAA AAC	AAC AAT TAC	GAA CCA AAC	CTA TTT AAA	ACT CCA CAA	AGG	429
His Lys Asn	Asn Asn Tyr	Glu Pro Asn	Leu Phe Lys	Thr Pro Gln	Arg	
	55		60	65		
AAA CCA TCT	TAT AAT CAG	CTG GCT TCA	ACT CCA ATA	ATA TTC AAA	GAG	477
Lys Pro Ser	Tyr Asn Gln	Leu Ala Ser	Thr Pro Ile	Ile Phe Lys	Glu	
70		75		80		
CAA GGG CTG	ACT CTG CCG	CTG TAC CAA	TCT CCT GTA	AAA GAA TTA	GAT	525
Gln Gly Leu	Thr Leu Pro	Leu Tyr Gln	Ser Pro Val	Lys Glu Leu	Asp	
85		90		95		
AAA TTC AAA	TTA GAC TTA	GGA AGG AAT	GTT CCC AAT	AGT AGA CAT	AAA	573
Lys Phe Lys	Leu Asp Leu	Gly Arg Asn	Val Pro Asn	Ser Arg His	Lys	
100		105		110	115	

6,033,857

63

64

-continued

AGT Ser	CTT Leu	CGC Arg	ACA Thr	GTG Val	AAA Lys	ACT Thr	AAA Lys	ATG Met	GAT Asp	CAA Gln	GCA Ala	GAT Asp	GAT Asp	GTT Val	TCC Ser	621
120125130																
TGT Cys	CCA Pro	CTT Leu	CTA Leu	AAT Asn	TCT Ser	TGT Cys	CTT Leu	AGT Ser	GAA Glu	AGT Ser	CCT Pro	GTT Val	GTT Val	CTA Leu	CAA Gln	669
135140145																
TGT Cys	ACA Thr	CAT His	GTA Val	ACA Thr	CCA Pro	CAA Gln	AGA Arg	GAT Asp	AAG Lys	TCA Ser	GTG Val	GTA Val	TGT Cys	GGG Gly	AGT Ser	717
150155160																
TTG Leu	TTT Phe	CAT His	ACA Thr	CCA Pro	AAG Lys	TTT Phe	GTG Val	AAG Lys	GGT Gly	CGT Arg	CAG Gln	ACA Thr	CCA Pro	AAA Lys	CAT His	765
165170175																
ATT Ile	TCT Ser	GAA Glu	AGT Ser	CTA Leu	GGA Gly	GCT Ala	GAG Glu	GTG Val	GAT Asp	CCT Pro	GAT Asp	ATG Met	TCT Ser	TGG Trp	TCA Ser	813
180185190195																
AGT Ser	TCT Ser	TTA Leu	GCT Ala	ACA Thr	CCA Pro	CCC Pro	ACC Thr	CTT Leu	AGT Ser	TCT Ser	ACT Thr	GTG Val	CTC Leu	ATA Ile	GTC Val	861
200205210																
AGA Arg	AAT Asn	GAA Glu	GAA Glu	GCA Ala	TCT Ser	GAA Glu	ACT Thr	GTA Val	TTT Phe	CCT Pro	CAT His	GAT Asp	ACT Thr	ACT Thr	GCT Ala	909
215220225																
AAT Asn	GTG Val	AAA Lys	AGC Ser	TAT Tyr	TTT Phe	TCC Ser	AAT Asn	CAT His	GAT Asp	GAA Glu	AGT Ser	CTG Leu	AAG Lys	AAA Lys	AAT Asn	957
230235240																
GAT Asp	AGA Arg	TTT Phe	ATC Ile	GCT Ala	TCT Ser	GTG Val	ACA Thr	GAC Asp	AGT Ser	GAA Glu	AAC Asn	ACA Thr	AAT Asn	CAA Gln	AGA Arg	1005
245250255																
GAA Glu	GCT Ala	GCA Ala	AGT Ser	CAT His	GGA Gly	TTT Phe	GGA Gly	AAA Lys	ACA Thr	TCA Ser	GGG Gly	AAT Asn	TCA Ser	TTT Phe	AAA Lys	1053
260265270275																
GTA Val	AAT Asn	AGC Ser	TGC Cys	AAA Lys	GAC Asp	CAC His	ATT Ile	GGA Gly	AAG Lys	TCA Ser	ATG Met	CCA Pro	AAT Asn	GTC Val	CTA Leu	1101
280285290																
GAA Glu	GAT Asp	GAA Glu	GTA Val	TAT Tyr	GAA Glu	ACA Thr	GTT Val	GTA Val	GAT Asp	ACC Thr	TCT Ser	GAA Glu	GAA Glu	GAT Asp	AGT Ser	1149
295300305																
TTT Phe	TCA Ser	TTA Leu	TGT Cys	TTT Phe	TCT Ser	AAA Lys	TGT Cys	AGA Arg	ACA Thr	AAA Lys	AAT Asn	CTA Leu	CAA Gln	AAA Lys	GTA Val	1197
310315320																
AGA Arg	ACT Thr	AGC Ser	AAG Lys	ACT Thr	AGG Arg	AAA Lys	AAA Lys	ATT Ile	TTC Phe	CAT His	GAA Glu	GCA Ala	AAC Asn	GCT Ala	GAT Asp	1245
325330335																
GAA Glu	TGT Cys	GAA Glu	AAA Lys	TCT Ser	AAA Lys	AAC Asn	CAA Gln	GTG Val	AAA Lys	GAA Glu	AAA Lys	TAC Tyr	TCA Ser	TTT Phe	GTA Val	1293
340345350355																
TCT Ser	GAA Glu	GTG Val	GAA Glu	CCA Pro	AAT Asn	GAT Asp	ACT Thr	GAT Asp	CCA Pro	TTA Leu	GAT Asp	TCA Ser	AAT Asn	GTA Val	GCA Ala	1341
360365370																
CAT His	CAG Gln	AAG Lys	CCC Pro	TTT Phe	GAG Glu	AGT Ser	GGA Gly	AGT Ser	GAC Asp	AAA Lys	ATC Ile	TCC Ser	AAG Lys	GAA Glu	GTT Val	1389
375380385																
GTA Val	CCG Pro	TCT Ser	TTG Leu	GCC Ala	TGT Cys	GAA Glu	TGG Trp	TCT Ser	CAA Gln	CTA Leu	ACC Thr	CTT Leu	TCA Ser	GGT Gly	CTA Leu	1437
390395400																
AAT Asn	GGA Gly	GCC Ala	CAG Gln	ATG Met	GAG Glu	AAA Lys	ATA Ile	CCC Pro	CTA Leu	TTG Leu	CAT His	ATT Ile	TCT Ser	TCA Ser	TGT Cys	1485
405410415																
GAC Asp	CAA Gln	AAT Asn	ATT Ile	TCA Ser	GAA Glu	AAA Lys	GAC Asp	CTA Leu	TTA Leu	GAC Asp	ACA Thr	GAG Glu	AAC Asn	AAA Lys	AGA Arg	1533
420425430435																

6,033,857

65

66

-continued

AAG AAA GAT TTT CTT ACT TCA GAG AAT TCT TTG CCA CGT ATT TCT AGC Lys Lys Asp Phe Leu Thr Ser Glu Asn Ser Leu Pro Arg Ile Ser Ser 440 445 450	1581
CTA CCA AAA TCA GAG AAG CCA TTA AAT GAG GAA ACA GTG GTA AAT AAG Leu Pro Lys Ser Glu Lys Pro Leu Asn Glu Glu Thr Val Val Asn Lys 455 460 465	1629
AGA GAT GAA GAG CAG CAT CTT GAA TCT CAT ACA GAC TGC ATT CTT GCA Arg Asp Glu Glu Gln His Leu Glu Ser His Thr Asp Cys Ile Leu Ala 470 475 480	1677
GTA AAG CAG GCA ATA TCT GGA ACT TCT CCA GTG GCT TCT TCA TTT CAG Val Lys Gln Ala Ile Ser Gly Thr Ser Pro Val Ala Ser Ser Phe Gln 485 490 495	1725
GGT ATC AAA AAG TCT ATA TTC AGA ATA AGA GAA TCA CCT AAA GAG ACT Gly Ile Lys Lys Ser Ile Phe Arg Ile Arg Glu Ser Pro Lys Glu Thr 500 505 510 515	1773
TTC AAT GCA AGT TTT TCA GGT CAT ATG ACT GAT CCA AAC TTT AAA AAA Phe Asn Ala Ser Phe Ser Gly His Met Thr Asp Pro Asn Phe Lys Lys 520 525 530	1821
GAA ACT GAA GCC TCT GAA AGT GGA CTG GAA ATA CAT ACT GTT TGC TCA Glu Thr Glu Ala Ser Glu Ser Gly Leu Glu Ile His Thr Val Cys Ser 535 540 545	1869
CAG AAG GAG GAC TCC TTA TGT CCA AAT TTA ATT GAT AAT GGA AGC TGG Gln Lys Glu Asp Ser Leu Cys Pro Asn Leu Ile Asp Asn Gly Ser Trp 550 555 560	1917
CCA GCC ACC ACC ACA CAG AAT TCT GTA GCT TTG AAG AAT GCA GGT TTA Pro Ala Thr Thr Thr Gln Asn Ser Val Ala Leu Lys Asn Ala Gly Leu 565 570 575	1965
ATA TCC ACT TTG AAA AAG AAA ACA AAT AAG TTT ATT TAT GCT ATA CAT Ile Ser Thr Leu Lys Lys Lys Thr Asn Lys Phe Ile Tyr Ala Ile His 580 585 590 595	2013
GAT GAA ACA TCT TAT AAA GGA AAA AAA ATA CCG AAA GAC CAA AAA TCA Asp Glu Thr Ser Tyr Lys Gly Lys Lys Ile Pro Lys Asp Gln Lys Ser 600 605 610	2061
GAA CTA ATT AAC TGT TCA GCC CAG TTT GAA GCA AAT GCT TTT GAA GCA Glu Leu Ile Asn Cys Ser Ala Gln Phe Glu Ala Asn Ala Phe Glu Ala 615 620 625	2109
CCA CTT ACA TTT GCA AAT GCT GAT TCA GGT TTA TTG CAT TCT TCT GTG Pro Leu Thr Phe Ala Asn Ala Asp Ser Gly Leu Leu His Ser Ser Val 630 635 640	2157
AAA AGA AGC TGT TCA CAG AAT GAT TCT GAA GAA CCA ACT TTG TCC TTA Lys Arg Ser Cys Ser Gln Asn Asp Ser Glu Glu Pro Thr Leu Ser Leu 645 650 655	2205
ACT AGC TCT TTT GGG ACA ATT CTG AGG AAA TGT TCT AGA AAT GAA ACA Thr Ser Ser Phe Gly Thr Ile Leu Arg Lys Cys Ser Arg Asn Glu Thr 660 665 670 675	2253
TGT TCT AAT AAT ACA GTA ATC TCT CAG GAT CTT GAT TAT AAA GAA GCA Cys Ser Asn Asn Thr Val Ile Ser Gln Asp Leu Asp Tyr Lys Glu Ala 680 685 690	2301
AAA TGT AAT AAG GAA AAA CTA CAG TTA TTT ATT ACC CCA GAA GCT GAT Lys Cys Asn Lys Glu Lys Leu Gln Leu Phe Ile Thr Pro Glu Ala Asp 695 700 705	2349
TCT CTG TCA TGC CTG CAG GAA GGA CAG TGT GAA AAT GAT CCA AAA AGC Ser Leu Ser Cys Leu Gln Glu Gly Gln Cys Glu Asn Asp Pro Lys Ser 710 715 720	2397
AAA AAA GTT TCA GAT ATA AAA GAA GAG GTC TTG GCT GCA GCA TGT CAC Lys Lys Val Ser Asp Ile Lys Glu Glu Val Leu Ala Ala Ala Cys His 725 730 735	2445
CCA GTA CAA CAT TCA AAA GTG GAA TAC AGT GAT ACT GAC TTT CAA TCC Pro Val Gln His Ser Lys Val Glu Tyr Ser Asp Thr Asp Phe Gln Ser	2493

A000490

6,033,857

67

68

-continued

740	745	750	755	
CAG AAA AGT CTT TTA TAT GAT CAT GAA AAT GCC AGC ACT CTT ATT TTA				2541
Gln Lys Ser Leu Leu Tyr Asp His Glu Asn Ala Ser Thr Leu Ile Leu				
	760	765	770	
ACT CCT ACT TCC AAG GAT GTT CTG TCA AAC CTA GTC ATG ATT TCT AGA				2589
Thr Pro Thr Ser Lys Asp Val Leu Ser Asn Leu Val Met Ile Ser Arg				
	775	780	785	
GGC AAA GAA TCA TAC AAA ATG TCA GAC AAG CTC AAA GGT AAC AAT TAT				2637
Gly Lys Glu Ser Tyr Lys Met Ser Asp Lys Leu Lys Gly Asn Asn Tyr				
	790	795	800	
GAA TCT GAT GTT GAA TTA ACC AAA AAT ATT CCC ATG GAA AAG AAT CAA				2685
Glu Ser Asp Val Glu Leu Thr Lys Asn Ile Pro Met Glu Lys Asn Gln				
	805	810	815	
GAT GTA TGT GCT TTA AAT GAA AAT TAT AAA AAC GTT GAG CTG TTG CCA				2733
Asp Val Cys Ala Leu Asn Glu Asn Tyr Lys Asn Val Glu Leu Leu Pro				
	820	825	830	835
CCT GAA AAA TAC ATG AGA GTA GCA TCA CCT TCA AGA AAG GTA CAA TTC				2781
Pro Glu Lys Tyr Met Arg Val Ala Ser Pro Ser Arg Lys Val Gln Phe				
	840	845	850	
AAC CAA AAC ACA AAT CTA AGA GTA ATC CAA AAA AAT CAA GAA GAA ACT				2829
Asn Gln Asn Thr Asn Leu Arg Val Ile Gln Lys Asn Gln Glu Glu Thr				
	855	860	865	
ACT TCA ATT TCA AAA ATA ACT GTC AAT CCA GAC TCT GAA GAA CTT TTC				2877
Thr Ser Ile Ser Lys Ile Thr Val Asn Pro Asp Ser Glu Glu Leu Phe				
	870	875	880	
TCA GAC AAT GAG AAT AAT TTT GTC TTC CAA GTA GCT AAT GAA AGG AAT				2925
Ser Asp Asn Glu Asn Asn Phe Val Phe Gln Val Ala Asn Glu Arg Asn				
	885	890	895	
AAT CTT GCT TTA GGA AAT ACT AAG GAA CTT CAT GAA ACA GAC TTG ACT				2973
Asn Leu Ala Leu Gly Asn Thr Lys Glu Leu His Glu Thr Asp Leu Thr				
	900	905	910	915
TGT GTA AAC GAA CCC ATT TTC AAG AAC TCT ACC ATG GTT TTA TAT GGA				3021
Cys Val Asn Glu Pro Ile Phe Lys Asn Ser Thr Met Val Leu Tyr Gly				
	920	925	930	
GAC ACA GGT GAT AAA CAA GCA ACC CAA GTG TCA ATT AAA AAA GAT TTG				3069
Asp Thr Gly Asp Lys Gln Ala Thr Gln Val Ser Ile Lys Lys Asp Leu				
	935	940	945	
GTT TAT GTT CTT GCA GAG GAG AAC AAA AAT AGT GTA AAG CAG CAT ATA				3117
Val Tyr Val Leu Ala Glu Glu Asn Lys Asn Ser Val Lys Gln His Ile				
	950	955	960	
AAA ATG ACT CTA GGT CAA GAT TTA AAA TCG GAC ATC TCC TTG AAT ATA				3165
Lys Met Thr Leu Gly Gln Asp Leu Lys Ser Asp Ile Ser Leu Asn Ile				
	965	970	975	
GAT AAA ATA CCA GAA AAA AAT AAT GAT TAC ATG AAC AAA TGG GCA GGA				3213
Asp Lys Ile Pro Glu Lys Asn Asn Asp Tyr Met Asn Lys Trp Ala Gly				
	980	985	990	995
CTC TTA GGT CCA ATT TCA AAT CAC AGT TTT GGA GGT AGC TTC AGA ACA				3261
Leu Leu Gly Pro Ile Ser Asn His Ser Phe Gly Gly Ser Phe Arg Thr				
	1000	1005	1010	
GCT TCA AAT AAG GAA ATC AAG CTC TCT GAA CAT AAC ATT AAG AAG AGC				3309
Ala Ser Asn Lys Glu Ile Lys Leu Ser Glu His Asn Ile Lys Lys Ser				
	1015	1020	1025	
AAA ATG TTC TTA AAA GAT ATT GAA GAA CAA TAT CCT ACT AGT TTA GCT				3357
Lys Met Phe Lys Asp Ile Glu Glu Gln Tyr Pro Thr Ser Leu Ala				
	1030	1035	1040	
TGT GTT GAA ATT GTA AAT ACC TTG GCA TTA GAT AAT CAA AAG AAA CTG				3405
Cys Val Glu Ile Val Asn Thr Leu Ala Leu Asp Asn Gln Lys Lys Leu				
	1045	1050	1055	
AGC AAG CCT CAG TCA ATT AAT ACT GTA TCT GCA CAT TTA CAG AGT AGT				3453

6,033,857

69

70

-continued

Ser Lys Pro Gln Ser Ile Asn Thr Val Ser Ala His Leu Gln Ser Ser	
1060 1065 1070 1075	
GTA GTT GTT TCT GAT TGT AAA AAT AGT CAT ATA ACC CCT CAG ATG TTA	3501
Val Val Val Ser Asp Cys Lys Asn Ser His Ile Thr Pro Gln Met Leu	
1080 1085 1090	
TTT TCC AAG CAG GAT TTT AAT TCA AAC CAT AAT TTA ACA CCT AGC CAA	3549
Phe Ser Lys Gln Asp Phe Asn Ser Asn His Asn Leu Thr Pro Ser Gln	
1095 1100 1105	
AAG GCA GAA ATT ACA GAA CTT TCT ACT ATA TTA GAA GAA TCA GGA AGT	3597
Lys Ala Glu Ile Thr Glu Leu Ser Thr Ile Leu Glu Glu Ser Gly Ser	
1110 1115 1120	
CAG TTT GAA TTT ACT CAG TTT AGA AAA CCA AGC TAC ATA TTG CAG AAG	3645
Gln Phe Glu Phe Thr Gln Phe Arg Lys Pro Ser Tyr Ile Leu Gln Lys	
1125 1130 1135	
AGT ACA TTT GAA GTG CCT GAA AAC CAG ATG ACT ATC TTA AAG ACC ACT	3693
Ser Thr Phe Glu Val Pro Glu Asn Gln Met Thr Ile Leu Lys Thr Thr	
1140 1145 1150 1155	
TCT GAG GAA TGC AGA GAT GCT GAT CTT CAT GTC ATA ATG AAT GCC CCA	3741
Ser Glu Glu Cys Arg Asp Ala Asp Leu His Val Ile Met Asn Ala Pro	
1160 1165 1170	
TCG ATT GGT CAG GTA GAC AGC AGC AAG CAA TTT GAA GGT ACA GTT GAA	3789
Ser Ile Gly Gln Val Asp Ser Ser Lys Gln Phe Glu Gly Thr Val Glu	
1175 1180 1185	
ATT AAA CGG AAG TTT GCT GGC CTG TTG AAA AAT GAC TGT AAC AAA AGT	3837
Ile Lys Arg Lys Phe Ala Gly Leu Leu Lys Asn Asp Cys Asn Lys Ser	
1190 1195 1200	
GCT TCT GGT TAT TTA ACA GAT GAA AAT GAA GTG GGG TTT AGG GGC TTT	3885
Ala Ser Gly Tyr Leu Thr Asp Glu Asn Glu Val Gly Phe Arg Gly Phe	
1205 1210 1215	
TAT TCT GCT CAT GGC ACA AAA CTG AAT GTT TCT ACT GAA GCT CTG CAA	3933
Tyr Ser Ala His Gly Thr Lys Leu Asn Val Ser Thr Glu Ala Leu Gln	
1220 1225 1230 1235	
AAA GCT GTG AAA CTG TTT AGT GAT ATT GAG AAT ATT AGT GAG GAA ACT	3981
Lys Ala Val Lys Leu Phe Ser Asp Ile Glu Asn Ile Ser Glu Glu Thr	
1240 1245 1250	
TCT GCA GAG GTA CAT CCA ATA AGT TTA TCT TCA AGT AAA TGT CAT GAT	4029
Ser Ala Glu Val His Pro Ile Ser Leu Ser Ser Ser Lys Cys His Asp	
1255 1260 1265	
TCT GTT GTT TCA ATG TTT AAG ATA GAA AAT CAT AAT GAT AAA ACT GTA	4077
Ser Val Val Ser Met Phe Lys Ile Glu Asn His Asn Asp Lys Thr Val	
1270 1275 1280	
AGT GAA AAA AAT AAT AAA TGC CAA CTG ATA TTA CAA AAT AAT ATT GAA	4125
Ser Glu Lys Asn Asn Lys Cys Gln Leu Ile Leu Gln Asn Asn Ile Glu	
1285 1290 1295	
ATG ACT ACT GGC ACT TTT GTT GAA GAA ATT ACT GAA AAT TAC AAG AGA	4173
Met Thr Thr Gly Thr Phe Val Glu Glu Ile Thr Glu Asn Tyr Lys Arg	
1300 1305 1310 1315	
AAT ACT GAA AAT GAA GAT AAC AAA TAT ACT GCT GCC AGT AGA AAT TCT	4221
Asn Thr Glu Asn Glu Asp Asn Lys Tyr Thr Ala Ala Ser Arg Asn Ser	
1320 1325 1330	
CAT AAC TTA GAA TTT GAT GGC AGT GAT TCA AGT AAA AAT GAT ACT GTT	4269
His Asn Leu Glu Phe Asp Gly Ser Asp Ser Ser Lys Asn Asp Thr Val	
1335 1340 1345	
TGT ATT CAT AAA GAT GAA ACG GAC TTG CTA TTT ACT GAT CAG CAC AAC	4317
Cys Ile His Lys Asp Glu Thr Asp Leu Leu Phe Thr Asp Gln His Asn	
1350 1355 1360	
ATA TGT CTT AAA TTA TCT GGC CAG TTT ATG AAG GAG GGA AAC ACT CAG	4365
Ile Cys Leu Lys Leu Ser Gly Gln Phe Met Lys Glu Gly Asn Thr Gln	
1365 1370 1375	

A000492

6,033,857

71

72

-continued

ATT AAA GAA GAT TTG TCA GAT TTA ACT TTT TTG GAA GTT GCG AAA GCT	4413
Ile Lys Glu Asp Leu Ser Asp Leu Thr Phe Leu Glu Val Ala Lys Ala	
1380 1385 1390 1395	
CAA GAA GCA TGT CAT GGT AAT ACT TCA AAT AAA GAA CAG TTA ACT GCT	4461
Gln Glu Ala Cys His Gly Asn Thr Ser Asn Lys Glu Gln Leu Thr Ala	
1400 1405 1410	
ACT AAA ACG GAG CAA AAT ATA AAA GAT TTT GAG ACT TCT GAT ACA TTT	4509
Thr Lys Thr Glu Gln Asn Ile Lys Asp Phe Glu Thr Ser Asp Thr Phe	
1415 1420 1425	
TTT CAG ACT GCA AGT GGG AAA AAT ATT AGT GTC GCC AAA GAG TCA TTT	4557
Phe Gln Thr Ala Ser Gly Lys Asn Ile Ser Val Ala Lys Glu Ser Phe	
1430 1435 1440	
AAT AAA ATT GTA AAT TTC TTT GAT CAG AAA CCA GAA GAA TTG CAT AAC	4605
Asn Lys Ile Val Asn Phe Phe Asp Gln Lys Pro Glu Glu Leu His Asn	
1445 1450 1455	
TTT TCC TTA AAT TCT GAA TTA CAT TCT GAC ATA AGA AAG AAC AAA ATG	4653
Phe Ser Leu Asn Ser Glu Leu His Ser Asp Ile Arg Lys Asn Lys Met	
1460 1465 1470 1475	
GAC ATT CTA AGT TAT GAG GAA ACA GAC ATA GTT AAA CAC AAA ATA CTG	4701
Asp Ile Leu Ser Tyr Glu Glu Thr Asp Ile Val Lys His Lys Ile Leu	
1480 1485 1490	
AAA GAA AGT GTC CCA GTT GGT ACT GGA AAT CAA CTA GTG ACC TTC CAG	4749
Lys Glu Ser Val Pro Val Gly Thr Gly Asn Gln Leu Val Thr Phe Gln	
1495 1500 1505	
GGA CAA CCC GAA CGT GAT GAA AAG ATC AAA GAA CCT ACT CTG TTG GGT	4797
Gly Gln Pro Glu Arg Asp Glu Lys Ile Lys Glu Pro Thr Leu Leu Gly	
1510 1515 1520	
TTT CAT ACA GCT AGC GGG AAA AAA GTT AAA ATT GCA AAG GAA TCT TTG	4845
Phe His Thr Ala Ser Gly Lys Lys Val Lys Ile Ala Lys Glu Ser Leu	
1525 1530 1535	
GAC AAA GTG AAA AAC CTT TTT GAT GAA AAA GAG CAA GGT ACT AGT GAA	4893
Asp Lys Val Lys Asn Leu Phe Asp Glu Lys Glu Gln Gly Thr Ser Glu	
1540 1545 1550 1555	
ATC ACC AGT TTT AGC CAT CAA TGG GCA AAG ACC CTA AAG TAC AGA GAG	4941
Ile Thr Ser Phe Ser His Gln Trp Ala Lys Thr Leu Lys Tyr Arg Glu	
1560 1565 1570	
GCC TGT AAA GAC CTT GAA TTA GCA TGT GAG ACC ATT GAG ATC ACA GCT	4989
Ala Cys Lys Asp Leu Glu Leu Ala Cys Glu Thr Ile Glu Ile Thr Ala	
1575 1580 1585	
GCC CCA AAG TGT AAA GAA ATG CAG AAT TCT CTC AAT AAT GAT AAA AAC	5037
Ala Pro Lys Cys Lys Glu Met Gln Asn Ser Leu Asn Asn Asp Lys Asn	
1590 1595 1600	
CTT GTT TCT ATT GAG ACT GTG GTG CCA CCT AAG CTC TTA AGT GAT AAT	5085
Leu Val Ser Ile Glu Thr Val Val Pro Pro Lys Leu Leu Ser Asp Asn	
1605 1610 1615	
TTA TGT AGA CAA ACT GAA AAT CTC AAA ACA TCA AAA AGT ATC TTT TTG	5133
Leu Cys Arg Gln Thr Glu Asn Leu Lys Thr Ser Lys Ser Ile Phe Leu	
1620 1625 1630 1635	
AAA GTT AAA GTA CAT GAA AAT GTA GAA AAA GAA ACA GCA AAA AGT CCT	5181
Lys Val Lys Val His Glu Asn Val Glu Lys Glu Thr Ala Lys Ser Pro	
1640 1645 1650	
GCA ACT TGT TAC ACA AAT CAG TCC CCT TAT TCA GTC ATT GAA AAT TCA	5229
Ala Thr Cys Tyr Thr Asn Gln Ser Pro Tyr Ser Val Ile Glu Asn Ser	
1655 1660 1665	
GCC TTA GCT TTT TAC ACA AGT TGT AGT AGA AAA ACT TCT GTG AGT CAG	5277
Ala Leu Ala Phe Tyr Thr Ser Cys Ser Arg Lys Thr Ser Val Ser Gln	
1670 1675 1680	
ACT TCA TTA CTT GAA GCA AAA AAA TGG CTT AGA GAA GGA ATA TTT GAT	5325
Thr Ser Leu Leu Glu Ala Lys Lys Trp Leu Arg Glu Gly Ile Phe Asp	
1685 1690 1695	

A000493

6,033,857

73

74

-continued

GGT CAA CCA GAA AGA ATA AAT ACT GCA GAT TAT GTA GGA AAT TAT TTG	5373
Gly Gln Pro Glu Arg Ile Asn Thr Ala Asp Tyr Val Gly Asn Tyr Leu	
1700 1705 1710 1715	
TAT GAA AAT AAT TCA AAC AGT ACT ATA GCT GAA AAT GAC AAA AAT CAT	5421
Tyr Glu Asn Asn Ser Asn Ser Thr Ile Ala Glu Asn Asp Lys Asn His	
1720 1725 1730	
CTC TCC GAA AAA CAA GAT ACT TAT TTA AGT AAC AGT AGC ATG TCT AAC	5469
Leu Ser Glu Lys Gln Asp Thr Tyr Leu Ser Asn Ser Ser Met Ser Asn	
1735 1740 1745	
AGC TAT TCC TAC CAT TCT GAT GAG GTA TAT AAT GAT TCA GGA TAT CTC	5517
Ser Tyr Ser Tyr His Ser Asp Glu Val Tyr Asn Asp Ser Gly Tyr Leu	
1750 1755 1760	
TCA AAA AAT AAA CTT GAT TCT GGT ATT GAG CCA GTA TTG AAG AAT GTT	5565
Ser Lys Asn Lys Leu Asp Ser Gly Ile Glu Pro Val Leu Lys Asn Val	
1765 1770 1775	
GAA GAT CAA AAA AAC ACT AGT TTT TCC AAA GTA ATA TCC AAT GTA AAA	5613
Glu Asp Gln Lys Asn Thr Ser Phe Ser Lys Val Ile Ser Asn Val Lys	
1780 1785 1790 1795	
GAT GCA AAT GCA TAC CCA CAA ACT GTA AAT GAA GAT ATT TGC GTT GAG	5661
Asp Ala Asn Ala Tyr Pro Gln Thr Val Asn Glu Asp Ile Cys Val Glu	
1800 1805 1810	
GAA CTT GTG ACT AGC TCT TCA CCC TGC AAA AAT AAA AAT GCA GCC ATT	5709
Glu Leu Val Thr Ser Ser Ser Pro Cys Lys Asn Lys Asn Ala Ala Ile	
1815 1820 1825	
AAA TTG TCC ATA TCT AAT AGT AAT AAT TTT GAG GTA GGG CCA CCT GCA	5757
Lys Leu Ser Ile Ser Asn Ser Asn Asn Phe Glu Val Gly Pro Pro Ala	
1830 1835 1840	
TTT AGG ATA GCC AGT GGT AAA ATC GTT TGT GTT TCA CAT GAA ACA ATT	5805
Phe Arg Ile Ala Ser Gly Lys Ile Val Cys Val Ser His Glu Thr Ile	
1845 1850 1855	
AAA AAA GTG AAA GAC ATA TTT ACA GAC AGT TTC AGT AAA GTA ATT AAG	5853
Lys Lys Val Lys Asp Ile Phe Thr Asp Ser Phe Ser Lys Val Ile Lys	
1860 1865 1870 1875	
GAA AAC AAC GAG AAT AAA TCA AAA ATT TGC CAA ACG AAA ATT ATG GCA	5901
Glu Asn Asn Glu Asn Lys Ser Lys Ile Cys Gln Thr Lys Ile Met Ala	
1880 1885 1890	
GGT TGT TAC GAG GCA TTG GAT GAT TCA GAG GAT ATT CTT CAT AAC TCT	5949
Gly Cys Tyr Glu Ala Leu Asp Asp Ser Glu Asp Ile Leu His Asn Ser	
1895 1900 1905	
CTA GAT AAT GAT GAA TGT AGC ACG CAT TCA CAT AAG GTT TTT GCT GAC	5997
Leu Asp Asn Asp Glu Cys Ser Thr His Ser His Lys Val Phe Ala Asp	
1910 1915 1920	
ATT CAG AGT GAA GAA ATT TTA CAA CAT AAC CAA AAT ATG TCT GGA TTG	6045
Ile Gln Ser Glu Glu Ile Leu Gln His Asn Gln Asn Met Ser Gly Leu	
1925 1930 1935	
GAG AAA GTT TCT AAA ATA TCA CCT TGT GAT GTT AGT TTG GAA ACT TCA	6093
Glu Lys Val Ser Lys Ile Ser Pro Cys Asp Val Ser Leu Glu Thr Ser	
1940 1945 1950 1955	
GAT ATA TGT AAA TGT AGT ATA GGG AAG CTT CAT AAG TCA GTC TCA TCT	6141
Asp Ile Cys Lys Cys Ser Ile Gly Lys Leu His Lys Ser Val Ser Ser	
1960 1965 1970	
GCA AAT ACT TGT GGG ATT TTT AGC ACA GCA AGT GGA AAA TCT GTC CAG	6189
Ala Asn Thr Cys Gly Ile Phe Ser Thr Ala Ser Gly Lys Ser Val Gln	
1975 1980 1985	
GTA TCA GAT GCT TCA TTA CAA AAC GCA AGA CAA GTG TTT TCT GAA ATA	6237
Val Ser Asp Ala Ser Leu Gln Asn Ala Arg Gln Val Phe Ser Glu Ile	
1990 1995 2000	
GAA GAT AGT ACC AAG CAA GTC TTT TCC AAA GTA TTG TTT AAA AGT AAC	6285
Glu Asp Ser Thr Lys Gln Val Phe Ser Lys Val Leu Phe Lys Ser Asn	

A000494

6,033,857

75

76

-continued

2005	2010	2015	
GAA CAT TCA GAC CAG CTC ACA AGA GAA GAA AAT ACT GCT ATA CGT ACT Glu His Ser Asp Gln Leu Thr Arg Glu Glu Asn Thr Ala Ile Arg Thr 2020 2025 2030 2035			6333
CCA GAA CAT TTA ATA TCC CAA AAA GGC TTT TCA TAT AAT GTG GTA AAT Pro Glu His Leu Ile Ser Gln Lys Gly Phe Ser Tyr Asn Val Val Asn 2040 2045 2050			6381
TCA TCT GCT TTC TCT GGA TTT AGT ACA GCA AGT GGA AAG CAA GTT TCC Ser Ser Ala Phe Ser Gly Phe Ser Thr Ala Ser Gly Lys Gln Val Ser 2055 2060 2065			6429
ATT TTA GAA AGT TCC TTA CAC AAA GTT AAG GGA GTG TTA GAG GAA TTT Ile Leu Glu Ser Ser Leu His Lys Val Lys Gly Val Leu Glu Glu Phe 2070 2075 2080			6477
GAT TTA ATC AGA ACT GAG CAT AGT CTT CAC TAT TCA CCT ACG TCT AGA Asp Leu Ile Arg Thr Glu His Ser Leu His Tyr Ser Pro Thr Ser Arg 2085 2090 2095			6525
CAA AAT GTA TCA AAA ATA CTT CCT CGT GTT GAT AAG AGA AAC CCA GAG Gln Asn Val Ser Lys Ile Leu Pro Arg Val Asp Lys Arg Asn Pro Glu 2100 2105 2110 2115			6573
CAC TGT GTA AAC TCA GAA ATG GAA AAA ACC TGC AGT AAA GAA TTT AAA His Cys Val Asn Ser Glu Met Glu Lys Thr Cys Ser Lys Glu Phe Lys 2120 2125 2130			6621
TTA TCA AAT AAC TTA AAT GTT GAA GGT GGT TCT TCA GAA AAT AAT CAC Leu Ser Asn Asn Leu Asn Val Glu Gly Gly Ser Ser Glu Asn Asn His 2135 2140 2145			6669
TCT ATT AAA GTT TCT CCA TAT CTC TCT CAA TTT CAA CAA GAC AAA CAA Ser Ile Lys Val Ser Pro Tyr Leu Ser Gln Phe Gln Gln Asp Lys Gln 2150 2155 2160			6717
CAG TTG GTA TTA GGA ACC AAA GTC TCA CTT GTT GAG AAC ATT CAT GTT Gln Leu Val Leu Gly Thr Lys Val Ser Leu Val Glu Asn Ile His Val 2165 2170 2175			6765
TTG GGA AAA GAA CAG GCT TCA CCT AAA AAC GTA AAA ATG GAA ATT GGT Leu Gly Lys Glu Gln Ala Ser Pro Lys Asn Val Lys Met Glu Ile Gly 2180 2185 2190 2195			6813
AAA ACT GAA ACT TTT TCT GAT GTT CCT GTG AAA ACA AAT ATA GAA GTT Lys Thr Glu Thr Phe Ser Asp Val Pro Val Lys Thr Asn Ile Glu Val 2200 2205 2210			6861
TGT TCT ACT TAC TCC AAA GAT TCA GAA AAC TAC TTT GAA ACA GAA GCA Cys Ser Thr Tyr Ser Lys Asp Ser Glu Asn Tyr Phe Glu Thr Glu Ala 2215 2220 2225			6909
GTA GAA ATT GCT AAA GCT TTT ATG GAA GAT GAT GAA CTG ACA GAT TCT Val Glu Ile Ala Lys Ala Phe Met Glu Asp Asp Glu Leu Thr Asp Ser 2230 2235 2240			6957
AAA CTG CCA AGT CAT GCC ACA CAT TCT CTT TTT ACA TGT CCC GAA AAT Lys Leu Pro Ser His Ala Thr His Ser Leu Phe Thr Cys Pro Glu Asn 2245 2250 2255			7005
GAG GAA ATG GTT TTG TCA AAT TCA AGA ATT GGA AAA AGA AGA GGA GAG Glu Glu Met Val Leu Ser Asn Ser Arg Ile Gly Lys Arg Arg Gly Glu 2260 2265 2270 2275			7053
CCC CTT ATC TTA GTG GGA GAA CCC TCA ATC AAA AGA AAC TTA TTA AAT Pro Leu Ile Leu Val Gly Glu Pro Ser Ile Lys Arg Asn Leu Leu Asn 2280 2285 2290			7101
GAA TTT GAC AGG ATA ATA GAA AAT CAA GAA AAA TCC TTA AAG GCT TCA Glu Phe Asp Arg Ile Glu Asn Gln Glu Lys Ser Leu Lys Ala Ser 2295 2300 2305			7149
AAA AGC ACT CCA GAT GGC ACA ATA AAA GAT CGA AGA TTG TTT ATG CAT Lys Ser Thr Pro Asp Gly Thr Ile Lys Asp Arg Arg Leu Phe Met His 2310 2315 2320			7197
CAT GTT TCT TTA GAG CCG ATT ACC TGT GTA CCC TTT CGC ACA ACT AAG			7245

77 6,033,857 78

-continued																	
His	Val	Ser	Leu	Glu	Pro	Ile	Thr	Cys	Val	Pro	Phe	Arg	Thr	Thr	Lys		
	2325						2330					2335					
GAA	CGT	CAA	GAG	ATA	CAG	AAT	CCA	AAT	TTT	ACC	GCA	CCT	GGT	CAA	GAA	7293	
Glu	Arg	Gln	Glu	Ile	Gln	Asn	Pro	Asn	Phe	Thr	Ala	Pro	Gly	Gln	Glu		
2340					2345					2350					2355		
TTT	CTG	TCT	AAA	TCT	CAT	TTG	TAT	GAA	CAT	CTG	ACT	TTG	GAA	AAA	TCT	7341	
Phe	Leu	Ser	Lys	Ser	His	Leu	Tyr	Glu	His	Leu	Thr	Leu	Glu	Lys	Ser		
			2360						2365					2370			
TCA	AGC	AAT	TTA	GCA	GTT	TCA	GGA	CAT	CCA	TTT	TAT	CAA	GTT	TCT	GCT	7389	
Ser	Ser	Asn	Leu	Ala	Val	Ser	Gly	His	Pro	Phe	Tyr	Gln	Val	Ser	Ala		
			2375					2380					2385				
ACA	AGA	AAT	GAA	AAA	ATG	AGA	CAC	TTG	ATT	ACT	ACA	GGC	AGA	CCA	ACC	7437	
Thr	Arg	Asn	Glu	Lys	Met	Arg	His	Leu	Ile	Thr	Thr	Gly	Arg	Pro	Thr		
			2390					2395					2400				
AAA	GTC	TTT	GTT	CCA	CCT	TTT	AAA	ACT	AAA	TCA	CAT	TTT	CAC	AGA	GTT	7485	
Lys	Val	Phe	Val	Pro	Pro	Phe	Lys	Thr	Lys	Ser	His	Phe	His	Arg	Val		
	2405					2410					2415						
GAA	CAG	TGT	GTT	AGG	AAT	ATT	AAC	TTG	GAG	GAA	AAC	AGA	CAA	AAG	CAA	7533	
Glu	Gln	Cys	Val	Arg	Asn	Ile	Asn	Leu	Glu	Glu	Asn	Arg	Gln	Lys	Gln		
2420					2425					2430					2435		
AAC	ATT	GAT	GGA	CAT	GGC	TCT	GAT	GAT	AGT	AAA	AAT	AAG	ATT	AAT	GAC	7581	
Asn	Ile	Asp	Gly	His	Gly	Ser	Asp	Asp	Ser	Lys	Asn	Lys	Ile	Asn	Asp		
			2440						2445					2450			
AAT	GAG	ATT	CAT	CAG	TTT	AAC	AAA	AAC	AAC	TCC	AAT	CAA	GCA	GCA	GCT	7629	
Asn	Glu	Ile	His	Gln	Phe	Asn	Lys	Asn	Asn	Ser	Asn	Gln	Ala	Ala	Ala		
			2455					2460					2465				
GTA	ACT	TTC	ACA	AAG	TGT	GAA	GAA	GAA	CCT	TTA	GAT	TTA	ATT	ACA	AGT	7677	
Val	Thr	Phe	Thr	Lys	Cys	Glu	Glu	Glu	Pro	Leu	Asp	Leu	Ile	Thr	Ser		
			2470					2475					2480				
CTT	CAG	AAT	GCC	AGA	GAT	ATA	CAG	GAT	ATG	CGA	ATT	AAG	AAG	AAA	CAA	7725	
Leu	Gln	Asn	Ala	Arg	Asp	Ile	Gln	Asp	Met	Arg	Ile	Lys	Lys	Lys	Gln		
	2485					2490					2495						
AGG	CAA	CGC	GTC	TTT	CCA	CAG	CCA	GGC	AGT	CTG	TAT	CTT	GCA	AAA	ACA	7773	
Arg	Gln	Arg	Val	Phe	Pro	Gln	Pro	Gly	Ser	Leu	Tyr	Leu	Ala	Lys	Thr		
2500					2505					2510					2515		
TCC	ACT	CTG	CCT	CGA	ATC	TCT	CTG	AAA	GCA	GCA	GTA	GGA	GGC	CAA	GTT	7821	
Ser	Thr	Leu	Pro	Arg	Ile	Ser	Leu	Lys	Ala	Ala	Val	Gly	Gly	Gln	Val		
					2520				2525					2530			
CCC	TCT	GCG	TGT	TCT	CAT	AAA	CAG	CTG	TAT	ACG	TAT	GGC	GTT	TCT	AAA	7869	
Pro	Ser	Ala	Cys	Ser	His	Lys	Gln	Leu	Tyr	Thr	Tyr	Gly	Val	Ser	Lys		
			2535					2540					2545				
CAT	TGC	ATA	AAA	ATT	AAC	AGC	AAA	AAT	GCA	GAG	TCT	TTT	CAG	TTT	CAC	7917	
His	Cys	Ile	Lys	Ile	Asn	Ser	Lys	Asn	Ala	Glu	Ser	Phe	Gln	Phe	His		
			2550					2555					2560				
ACT	GAA	GAT	TAT	TTT	GGT	AAG	GAA	AGT	TTA	TGG	ACT	GGA	AAA	GGA	ATA	7965	
Thr	Glu	Asp	Tyr	Phe	Gly	Lys	Glu	Ser	Leu	Trp	Thr	Gly	Lys	Gly	Ile		
	2565					2570					2575						
CAG	TTG	GCT	GAT	GGT	GGA	TGG	CTC	ATA	CCC	TCC	AAT	GAT	GGA	AAG	GCT	8013	
Gln	Leu	Ala	Asp	Gly	Gly	Trp	Leu	Ile	Pro	Ser	Asn	Asp	Gly	Lys	Ala		
2580					2585					2590					2595		
GGA	AAA	GAA	GAA	TTT	TAT	AGG	GCT	CTG	TGT	GAC	ACT	CCA	GGT	GTG	GAT	8061	
Gly	Lys	Glu	Glu	Phe	Tyr	Arg	Ala	Leu	Cys	Asp	Thr	Pro	Gly	Val	Asp		
			2600						2605					2610			
CCA	AAG	CTT	ATT	TCT	AGA	ATT	TGG	GTT	TAT	AAT	CAC	TAT	AGA	TGG	ATC	8109	
Pro	Lys	Leu	Ile	Ser	Arg	Ile	Trp	Val	Tyr	Asn	His	Tyr	Arg	Trp	Ile		
			2615					2620					2625				
ATA	TGG	AAA	CTG	GCA	GCT	ATG	GAA	TGT	GCC	TTT	CCT	AAG	GAA	TTT	GCT	8157	
Ile	Trp	Lys	Leu	Ala	Ala	Met	Glu	Cys	Ala	Phe	Pro	Lys	Glu	Phe	Ala		
			2630					2635					2640				

6,033,857

79

80

-continued

AAT AGA TGC CTA AGC CCA GAA AGG GTG CTT CTT CAA CTA AAA TAC AGA Asn Arg Cys Leu Ser Pro Glu Arg Val Leu Leu Gln Leu Lys Tyr Arg 2645 2650 2655	8205
TAT GAT ACG GAA ATT GAT AGA AGC AGA AGA TCG GCT ATA AAA AAG ATA Tyr Asp Thr Glu Ile Asp Arg Ser Arg Arg Ser Ala Ile Lys Lys Ile 2660 2665 2670 2675	8253
ATG GAA AGG GAT GAC ACA GCT GCA AAA ACA CTT GTT CTC TGT GTT TCT Met Glu Arg Asp Asp Thr Ala Ala Lys Thr Leu Val Leu Cys Val Ser 2680 2685 2690	8301
GAC ATA ATT TCA TTG AGC GCA AAT ATA TCT GAA ACT TCT AGC AAT AAA Asp Ile Ile Ser Leu Ser Ala Asn Ile Ser Glu Thr Ser Ser Asn Lys 2695 2700 2705	8349
ACT AGT AGT GCA GAT ACC CAA AAA GTG GCC ATT ATT GAA CTT ACA GAT Thr Ser Ser Ala Asp Thr Gln Lys Val Ala Ile Ile Glu Leu Thr Asp 2710 2715 2720	8397
GGG TGG TAT GCT GTT AAG GCC CAG TTA GAT CCT CCC CTC TTA GCT GTC Gly Trp Tyr Ala Val Lys Ala Gln Leu Asp Pro Pro Leu Leu Ala Val 2725 2730 2735	8445
TTA AAG AAT GGC AGA CTG ACA GTT GGT CAG AAG ATT ATT CTT CAT GGA Leu Lys Asn Gly Arg Leu Thr Val Gly Gln Lys Ile Ile Leu His Gly 2740 2745 2750 2755	8493
GCA GAA CTG GTG GGC TCT CCT GAT GCC TGT ACA CCT CTT GAA GCC CCA Ala Glu Leu Val Gly Ser Pro Asp Ala Cys Thr Pro Leu Glu Ala Pro 2760 2765 2770	8541
GAA TCT CTT ATG TTA AAG ATT TCT GCT AAC AGT ACT CGG CCT GCT CGC Glu Ser Leu Met Leu Lys Ile Ser Ala Asn Ser Thr Arg Pro Ala Arg 2775 2780 2785	8589
TGG TAT ACC AAA CTT GGA TTC TTT CCT GAC CCT AGA CCT TTT CCT CTG Trp Tyr Thr Lys Leu Gly Phe Phe Pro Asp Pro Arg Pro Phe Pro Leu 2790 2795 2800	8637
CCC TTA TCA TCG CTT TTC AGT GAT GGA GGA AAT GTT GGT TGT GTT GAT Pro Leu Ser Ser Leu Phe Ser Asp Gly Gly Asn Val Gly Cys Val Asp 2805 2810 2815	8685
GTA ATT ATT CAA AGA GCA TAC CCT ATA CAG TGG ATG GAG AAG ACA TCA Val Ile Ile Gln Arg Ala Tyr Pro Ile Gln Trp Met Glu Lys Thr Ser 2820 2825 2830 2835	8733
TCT GGA TTA TAC ATA TTT CGC AAT GAA AGA GAG GAA GAA AAG GAA GCA Ser Gly Leu Tyr Ile Phe Arg Asn Glu Arg Glu Glu Glu Lys Glu Ala 2840 2845 2850	8781
GCA AAA TAT GTG GAG GCC CAA CAA AAG AGA CTA GAA GCC TTA TTC ACT Ala Lys Tyr Val Glu Ala Gln Gln Lys Arg Leu Glu Ala Leu Phe Thr 2855 2860 2865	8829
AAA ATT CAG GAG GAA TTT GAA GAA CAT GAA GAA AAC ACA ACA AAA CCA Lys Ile Gln Glu Glu Phe Glu Glu His Glu Glu Asn Thr Thr Lys Pro 2870 2875 2880	8877
TAT TTA CCA TCA CGT GCA CTA ACA AGA CAG CAA GTT CGT GCT TTG CAA Tyr Leu Pro Ser Arg Ala Leu Thr Arg Gln Gln Val Arg Ala Leu Gln 2885 2890 2895	8925
GAT GGT GCA GAG CTT TAT GAA GCA GTG AAG AAT GCA GCA GAC CCA GCT Asp Gly Ala Glu Leu Tyr Glu Ala Val Lys Asn Ala Ala Asp Pro Ala 2900 2905 2910 2915	8973
TAC CTT GAG GGT TAT TTC AGT GAA GAG CAG TTA AGA GCC TTG AAT AAT Tyr Leu Glu Gly Tyr Phe Ser Glu Glu Gln Leu Arg Ala Leu Asn Asn 2920 2925 2930	9021
CAC AGG CAA ATG TTG AAT GAT AAG AAA CAA GCT CAG ATC CAG TTG GAA His Arg Gln Met Leu Asn Asp Lys Lys Gln Ala Gln Ile Gln Leu Glu 2935 2940 2945	9069
ATT AGG AAG GCC ATG GAA TCT GCT GAA CAA AAG GAA CAA GGT TTA TCA Ile Arg Lys Ala Met Glu Ser Ala Glu Gln Lys Glu Gln Gly Leu Ser 2950 2955 2960	9117

A000497

6,033,857

81

82

-continued

AGG GAT GTC ACA ACC GTG TGG AAG TTG CGT ATT GTA AGC TAT TCA AAA 9165 Arg Asp Val Thr Thr Val Trp Lys Leu Arg Ile Val Ser Tyr Ser Lys 2965 2970 2975
AAA GAA AAA GAT TCA GTT ATA CTG AGT ATT TGG CGT CCA TCA TCA GAT 9213 Lys Glu Lys Asp Ser Val Ile Leu Ser Ile Trp Arg Pro Ser Ser Asp 2980 2985 2990 2995
TTA TAT TCT CTG TTA ACA GAA GGA AAG AGA TAC AGA ATT TAT CAT CTT 9261 Leu Tyr Ser Leu Leu Thr Glu Gly Lys Arg Tyr Arg Ile Tyr His Leu 3000 3005 3010
GCA ACT TCA AAA TCT AAA AGT AAA TCT GAA AGA GCT AAC ATA CAG TTA 9309 Ala Thr Ser Lys Ser Lys Ser Lys Ser Glu Arg Ala Asn Ile Gln Leu 3015 3020 3025
GCA GCG ACA AAA AAA ACT CAG TAT CAA CAA CTA CCG GTT TCA GAT GAA 9357 Ala Ala Thr Lys Lys Thr Gln Tyr Gln Gln Leu Pro Val Ser Asp Glu 3030 3035 3040
ATT TTA TTT CAG ATT TAC CAG CCA CGG GAG CCC CTT CAC TTC AGC AAA 9405 Ile Leu Phe Gln Ile Tyr Gln Pro Arg Glu Pro Leu His Phe Ser Lys 3045 3050 3055
TTT TTA GAT CCA GAC TTT CAG CCA TCT TGT TCT GAG GTG GAC CTA ATA 9453 Phe Leu Asp Pro Asp Phe Gln Pro Ser Cys Ser Glu Val Asp Leu Ile 3060 3065 3070 3075
GGA TTT GTC GTT TCT GTT GTG AAA AAA ACA GGA CTT GCC CCT TTC GTC 9501 Gly Phe Val Val Ser Val Val Lys Lys Thr Gly Leu Ala Pro Phe Val 3080 3085 3090
TAT TTG TCA GAC GAA TGT TAC AAT TTA CTG GCA ATA AAG TTT TGG ATA 9549 Tyr Leu Ser Asp Glu Cys Tyr Asn Leu Leu Ala Ile Lys Phe Trp Ile 3095 3100 3105
GAC CTT AAT GAG GAC ATT ATT AAG CCT CAT ATG TTA ATT GCT GCA AGC 9597 Asp Leu Asn Glu Asp Ile Ile Lys Pro His Met Leu Ile Ala Ala Ser 3110 3115 3120
AAC CTC CAG TGG CGA CCA GAA TCC AAA TCA GGC CTT CTT ACT TTA TTT 9645 Asn Leu Gln Trp Arg Pro Glu Ser Lys Ser Gly Leu Leu Thr Leu Phe 3125 3130 3135
GCT GGA GAT TTT TCT GTG TTT TCT GCT AGT CCA AAA GAG GGC CAC TTT 9693 Ala Gly Asp Phe Ser Val Phe Ser Ala Ser Pro Lys Glu Gly His Phe 3140 3145 3150 3155
CAA GAG ACA TTC AAC AAA ATG AAA AAT ACT GTT GAG AAT ATT GAC ATA 9741 Gln Glu Thr Phe Asn Lys Met Lys Asn Thr Val Glu Asn Ile Asp Ile 3160 3165 3170
CTT TGC AAT GAA GCA GAA AAC AAG CTT ATG CAT ATA CTG CAT GCA AAT 9789 Leu Cys Asn Glu Ala Glu Asn Lys Leu Met His Ile Leu His Ala Asn 3175 3180 3185
GAT CCC AAG TGG TCC ACC CCA ACT AAA GAC TGT ACT TCA GGG CCG TAC 9837 Asp Pro Lys Trp Ser Thr Pro Thr Lys Asp Cys Thr Ser Gly Pro Tyr 3190 3195 3200
ACT GCT CAA ATC ATT CCT GGT ACA GGA AAC AAG CTT CTG ATG TCT TCT 9885 Thr Ala Gln Ile Ile Pro Gly Thr Gly Asn Lys Leu Leu Met Ser Ser 3205 3210 3215
CCT AAT TGT GAG ATA TAT TAT CAA AGT CCT TTA TCA CTT TGT ATG GCC 9933 Pro Asn Cys Glu Ile Tyr Tyr Gln Ser Pro Leu Ser Leu Cys Met Ala 3220 3225 3230 3235
AAA AGG AAG TCT GTT TCC ACA CCT GTC TCA GCC CAG ATG ACT TCA AAG 9981 Lys Arg Lys Ser Val Ser Thr Pro Val Ser Ala Gln Met Thr Ser Lys 3240 3245 3250
TCT TGT AAA GGG GAG AAA GAG ATT GAT GAC CAA AAG AAC TGC AAA AAG 10029 Ser Cys Lys Gly Glu Lys Glu Ile Asp Asp Gln Lys Asn Cys Lys Lys 3255 3260 3265
AGA AGA GCC TTG GAT TTC TTG AGT AGA CTG CCT TTA CCT CCA CCT GTT 10077 Arg Arg Ala Leu Asp Phe Leu Ser Arg Leu Pro Leu Pro Pro Pro Val

A000498

6,033,857

83

84

-continued

3270	3275	3280	
AGT CCC ATT TGT ACA TTT GTT TCT CCG GCT GCA CAG AAG GCA TTT CAG			10125
Ser Pro Ile Cys Thr Phe Val Ser Pro Ala Ala Gln Lys Ala Phe Gln			
3285	3290	3295	
CCA CCA AGG AGT TGT GGC ACC AAA TAC GAA ACA CCC ATA AAG AAA AAA			10173
Pro Pro Arg Ser Cys Gly Thr Lys Tyr Glu Thr Pro Ile Lys Lys Lys			
3300	3305	3310	3315
GAA CTG AAT TCT CCT CAG ATG ACT CCA TTT AAA AAA TTC AAT GAA ATT			10221
Glu Leu Asn Ser Pro Gln Met Thr Pro Phe Lys Lys Phe Asn Glu Ile			
3320	3325	3330	
TCT CTT TTG GAA AGT AAT TCA ATA GCT GAC GAA GAA CTT GCA TTG ATA			10269
Ser Leu Leu Glu Ser Asn Ser Ile Ala Asp Glu Glu Leu Ala Leu Ile			
3335	3340	3345	
AAT ACC CAA GCT CTT TTG TCT GGT TCA ACA GGA GAA AAA CAA TTT ATA			10317
Asn Thr Gln Ala Leu Leu Ser Gly Ser Thr Gly Glu Lys Gln Phe Ile			
3350	3355	3360	
TCT GTC AGT GAA TCC ACT AGG ACT GCT CCC ACC AGT TCA GAA GAT TAT			10365
Ser Val Ser Glu Ser Thr Arg Thr Ala Pro Thr Ser Ser Glu Asp Tyr			
3365	3370	3375	
CTC AGA CTG AAA CGA CGT TGT ACT ACA TCT CTG ATC AAA GAA CAG GAG			10413
Leu Arg Leu Lys Arg Arg Cys Thr Thr Ser Leu Ile Lys Glu Gln Glu			
3380	3385	3390	3395
AGT TCC CAG GCC AGT ACG GAA GAA TGT GAG AAA AAT AAG CAG GAC ACA			10461
Ser Ser Gln Ala Ser Thr Glu Glu Cys Glu Lys Asn Lys Gln Asp Thr			
3400	3405	3410	
ATT ACA ACT AAA AAA TAT ATC TAAGCATTG CAAAGGCGAC AATAAATTAT			10512
Ile Thr Thr Lys Lys Tyr Ile			
3415			
TGACGCTTAA CCTTTCAGT TTATAAGACT GGAATATAAT TTCAAACCAC ACATTAGTAC			10572
TTATGTTGCA CAATGAGAAA AGAAATTAGT TTCAAATTTA CCTCAGCGTT TGTGTATCGG			10632
GCAAAAATCG TTTTGCCCGA TTCCGTATTG GTATACTTTT GCTTCAGTTG CATATCTTAA			10692
AACTAAATGT AATTATTATA CTAATCAAGA AAAACATCTT TGGCTGAGCT CGGTGGCTCA			10752
TGCCTGTAAT CCCAACACTT TGAGAAGCTG AGGTGGGAGG AGTGCTTGAG GCCAGGAGTT			10812
CAAGACCAGC CTGGGCAACA TAGGGAGACC CCCATCTTTA CGAAGAAAAA AAAAAAGGGG			10872
AAAAGAAAAT CTTTAAATC TTTGGATTG ATCACTACAA GTATTATTTT ACAAGTGAAA			10932
TAAACATACC ATTTTCTTTT AGATTGTGTC ATTAATGGA ATGAGGCTC TTAGTACAGT			10992
TATTTTGATG CAGATAATTC CTTTLAGTTT AGCTACTATT TTAGGGGATT TTTTGTAGAG			11052
GTAACCTACT ATGAAATAGT TCTCCTTAAT GCAAATATGT TGGTCTGCT ATAGTTCCAT			11112
CCTGTTCAAA AGTCAGGATG AATATGAAGA GTGGTGTTTC CTTTGAGCA ATTCTTCATC			11172
CTTAAGTCAG CATGATTATA AGAAAAATAG AACCTCAGT GAACTCTAA TTCCTTTTTA			11232
CTATTCAGT GTGATCTCTG AAATTAAATT ACTTCAACTA AAAATTCAAA TACTTTAAAT			11292
CAGAAGATTT CATAGTTAAT TTATTTTTTT TTTCAACAAA ATGTCATCC AAACCAAAC			11352
TTGAGAAAAT ATCTTGCTTT CAAATTGACA CTA			11385

- (2) INFORMATION FOR SEQ ID NO:2:
- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 3418 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: protein
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

6,033,857

85

86

-continued

Met	Pro	Ile	Gly	Ser	Lys	Glu	Arg	Pro	Thr	Phe	Phe	Glu	Ile	Phe	Lys	1	5	10	15
Thr	Arg	Cys	Asn	Lys	Ala	Asp	Leu	Gly	Pro	Ile	Ser	Leu	Asn	Trp	Phe	20	25	30	
Glu	Glu	Leu	Ser	Ser	Glu	Ala	Pro	Pro	Tyr	Asn	Ser	Glu	Pro	Ala	Glu	35	40	45	
Glu	Ser	Glu	His	Lys	Asn	Asn	Asn	Tyr	Glu	Pro	Asn	Leu	Phe	Lys	Thr	50	55	60	
Pro	Gln	Arg	Lys	Pro	Ser	Tyr	Asn	Gln	Leu	Ala	Ser	Thr	Pro	Ile	Ile	65	70	75	80
Phe	Lys	Glu	Gln	Gly	Leu	Thr	Leu	Pro	Leu	Tyr	Gln	Ser	Pro	Val	Lys	85	90	95	
Glu	Leu	Asp	Lys	Phe	Lys	Leu	Asp	Leu	Gly	Arg	Asn	Val	Pro	Asn	Ser	100	105	110	
Arg	His	Lys	Ser	Leu	Arg	Thr	Val	Lys	Thr	Lys	Met	Asp	Gln	Ala	Asp	115	120	125	
Asp	Val	Ser	Cys	Pro	Leu	Leu	Asn	Ser	Cys	Leu	Ser	Glu	Ser	Pro	Val	130	135	140	
Val	Leu	Gln	Cys	Thr	His	Val	Thr	Pro	Gln	Arg	Asp	Lys	Ser	Val	Val	145	150	155	160
Cys	Gly	Ser	Leu	Phe	His	Thr	Pro	Lys	Phe	Val	Lys	Gly	Arg	Gln	Thr	165	170	175	
Pro	Lys	His	Ile	Ser	Glu	Ser	Leu	Gly	Ala	Glu	Val	Asp	Pro	Asp	Met	180	185	190	
Ser	Trp	Ser	Ser	Ser	Leu	Ala	Thr	Pro	Pro	Thr	Leu	Ser	Ser	Thr	Val	195	200	205	
Leu	Ile	Val	Arg	Asn	Glu	Glu	Ala	Ser	Glu	Thr	Val	Phe	Pro	His	Asp	210	215	220	
Thr	Thr	Ala	Asn	Val	Lys	Ser	Tyr	Phe	Ser	Asn	His	Asp	Glu	Ser	Leu	225	230	235	240
Lys	Lys	Asn	Asp	Arg	Phe	Ile	Ala	Ser	Val	Thr	Asp	Ser	Glu	Asn	Thr	245	250	255	
Asn	Gln	Arg	Glu	Ala	Ala	Ser	His	Gly	Phe	Gly	Lys	Thr	Ser	Gly	Asn	260	265	270	
Ser	Phe	Lys	Val	Asn	Ser	Cys	Lys	Asp	His	Ile	Gly	Lys	Ser	Met	Pro	275	280	285	
Asn	Val	Leu	Glu	Asp	Glu	Val	Tyr	Glu	Thr	Val	Val	Asp	Thr	Ser	Glu	290	295	300	
Glu	Asp	Ser	Phe	Ser	Leu	Cys	Phe	Ser	Lys	Cys	Arg	Thr	Lys	Asn	Leu	305	310	315	320
Gln	Lys	Val	Arg	Thr	Ser	Lys	Thr	Arg	Lys	Lys	Ile	Phe	His	Glu	Ala	325	330	335	
Asn	Ala	Asp	Glu	Cys	Glu	Lys	Ser	Lys	Asn	Gln	Val	Lys	Glu	Lys	Tyr	340	345	350	
Ser	Phe	Val	Ser	Glu	Val	Glu	Pro	Asn	Asp	Thr	Asp	Pro	Leu	Asp	Ser	355	360	365	
Asn	Val	Ala	His	Gln	Lys	Pro	Phe	Glu	Ser	Gly	Ser	Asp	Lys	Ile	Ser	370	375	380	
Lys	Glu	Val	Val	Pro	Ser	Leu	Ala	Cys	Glu	Trp	Ser	Gln	Leu	Thr	Leu	385	390	395	400
Ser	Gly	Leu	Asn	Gly	Ala	Gln	Met	Glu	Lys	Ile	Pro	Leu	Leu	His	Ile	405	410	415	

A000500

6,033,857

87

88

-continued

Ser	Ser	Cys	Asp	Gln	Asn	Ile	Ser	Glu	Lys	Asp	Leu	Leu	Asp	Thr	Glu
			420					425					430		
Asn	Lys	Arg	Lys	Lys	Asp	Phe	Leu	Thr	Ser	Glu	Asn	Ser	Leu	Pro	Arg
		435					440					445			
Ile	Ser	Ser	Leu	Pro	Lys	Ser	Glu	Lys	Pro	Leu	Asn	Glu	Glu	Thr	Val
	450				455						460				
Val	Asn	Lys	Arg	Asp	Glu	Gln	His	Leu	Glu	Ser	His	Thr	Asp	Cys	
465					470				475					480	
Ile	Leu	Ala	Val	Lys	Gln	Ala	Ile	Ser	Gly	Thr	Ser	Pro	Val	Ala	Ser
				485					490					495	
Ser	Phe	Gln	Gly	Ile	Lys	Lys	Ser	Ile	Phe	Arg	Ile	Arg	Glu	Ser	Pro
			500					505					510		
Lys	Glu	Thr	Phe	Asn	Ala	Ser	Phe	Ser	Gly	His	Met	Thr	Asp	Pro	Asn
		515					520					525			
Phe	Lys	Lys	Glu	Thr	Glu	Ala	Ser	Glu	Ser	Gly	Leu	Glu	Ile	His	Thr
	530					535					540				
Val	Cys	Ser	Gln	Lys	Glu	Asp	Ser	Leu	Cys	Pro	Asn	Leu	Ile	Asp	Asn
545				550					555					560	
Gly	Ser	Trp	Pro	Ala	Thr	Thr	Thr	Gln	Asn	Ser	Val	Ala	Leu	Lys	Asn
			565					570						575	
Ala	Gly	Leu	Ile	Ser	Thr	Leu	Lys	Lys	Lys	Thr	Asn	Lys	Phe	Ile	Tyr
			580				585						590		
Ala	Ile	His	Asp	Glu	Thr	Phe	Tyr	Lys	Gly	Lys	Lys	Ile	Pro	Lys	Asp
		595					600					605			
Gln	Lys	Ser	Glu	Leu	Ile	Asn	Cys	Ser	Ala	Gln	Phe	Glu	Ala	Asn	Ala
	610					615					620				
Phe	Glu	Ala	Pro	Leu	Thr	Phe	Ala	Asn	Ala	Asp	Ser	Gly	Leu	Leu	His
625				630					635					640	
Ser	Ser	Val	Lys	Arg	Ser	Cys	Ser	Gln	Asn	Asp	Ser	Glu	Glu	Pro	Thr
			645					650						655	
Leu	Ser	Leu	Thr	Ser	Ser	Phe	Gly	Thr	Ile	Leu	Arg	Lys	Cys	Ser	Arg
		660					665						670		
Asn	Glu	Thr	Cys	Ser	Asn	Asn	Thr	Val	Ile	Ser	Gln	Asp	Leu	Asp	Tyr
		675					680					685			
Lys	Glu	Ala	Lys	Cys	Asn	Lys	Glu	Lys	Leu	Gln	Leu	Phe	Ile	Thr	Pro
		690				695					700				
Glu	Ala	Asp	Ser	Leu	Ser	Cys	Leu	Gln	Glu	Gly	Gln	Cys	Glu	Asn	Asp
705				710					715					720	
Pro	Lys	Ser	Lys	Lys	Val	Ser	Asp	Ile	Lys	Glu	Glu	Val	Leu	Ala	Ala
			725					730					735		
Ala	Cys	His	Pro	Val	Gln	His	Ser	Lys	Val	Glu	Tyr	Ser	Asp	Thr	Asp
			740				745						750		
Phe	Gln	Ser	Gln	Lys	Ser	Leu	Leu	Tyr	Asp	His	Glu	Asn	Ala	Ser	Thr
		755					760					765			
Leu	Ile	Leu	Thr	Pro	Thr	Ser	Lys	Asp	Val	Leu	Ser	Asn	Leu	Val	Met
		770				775						780			
Ile	Ser	Arg	Gly	Lys	Glu	Ser	Tyr	Lys	Met	Ser	Asp	Lys	Leu	Lys	Gly
785				790					795					800	
Asn	Asn	Tyr	Glu	Ser	Asp	Val	Glu	Leu	Thr	Lys	Asn	Ile	Pro	Met	Glu
			805					810					815		
Lys	Asn	Gln	Asp	Val	Cys	Ala	Leu	Asn	Glu	Asn	Tyr	Lys	Asn	Val	Glu
			820				825						830		
Leu	Leu	Pro	Pro	Glu	Lys	Tyr	Met	Arg	Val	Ala	Ser	Pro	Ser	Arg	Lys

6,033,857

89

90

-continued

835						840						845											
Val	Gln	Phe	Asn	Gln	Asn	Thr	Asn	Leu	Arg	Val	Ile	Gln	Lys	Asn	Gln								
850						855						860											
Glu	Glu	Thr	Thr	Ser	Ile	Ser	Lys	Ile	Thr	Val	Asn	Pro	Asp	Ser	Glu								
865						870						875						880					
Glu	Leu	Phe	Ser	Asp	Asn	Glu	Asn	Asn	Phe	Val	Phe	Gln	Val	Ala	Asn								
						885						890						895					
Glu	Arg	Asn	Asn	Leu	Ala	Leu	Gly	Asn	Thr	Lys	Glu	Leu	His	Glu	Thr								
900						905						910											
Asp	Leu	Thr	Cys	Val	Asn	Glu	Pro	Ile	Phe	Lys	Asn	Ser	Thr	Met	Val								
915						920						925											
Leu	Tyr	Gly	Asp	Thr	Gly	Asp	Lys	Gln	Ala	Thr	Gln	Val	Ser	Ile	Lys								
930						935						940											
Lys	Asp	Leu	Val	Tyr	Val	Leu	Ala	Glu	Glu	Asn	Lys	Asn	Ser	Val	Lys								
945						950						955						960					
Gln	His	Ile	Lys	Met	Thr	Leu	Gly	Gln	Asp	Leu	Lys	Ser	Asp	Ile	Ser								
						965						970						975					
Leu	Asn	Ile	Asp	Lys	Ile	Pro	Glu	Lys	Asn	Asn	Asp	Tyr	Met	Asn	Lys								
980						985						990											
Trp	Ala	Gly	Leu	Leu	Gly	Pro	Ile	Ser	Asn	His	Ser	Phe	Gly	Gly	Ser								
995						1000						1005											
Phe	Arg	Thr	Ala	Ser	Asn	Lys	Glu	Ile	Lys	Leu	Ser	Glu	His	Asn	Ile								
1010						1015						1020											
Lys	Lys	Ser	Lys	Met	Phe	Phe	Lys	Asp	Ile	Glu	Glu	Gln	Tyr	Pro	Thr								
1025						1030						1035						1040					
Ser	Leu	Ala	Cys	Val	Glu	Ile	Val	Asn	Thr	Leu	Ala	Leu	Asp	Asn	Gln								
						1045						1050						1055					
Lys	Lys	Leu	Ser	Lys	Pro	Gln	Ser	Ile	Asn	Thr	Val	Ser	Ala	His	Leu								
1060						1065						1070											
Gln	Ser	Ser	Val	Val	Val	Ser	Asp	Cys	Lys	Asn	Ser	His	Ile	Thr	Pro								
1075						1080						1085											
Gln	Met	Leu	Phe	Ser	Lys	Gln	Asp	Phe	Asn	Ser	Asn	His	Asn	Leu	Thr								
1090						1095						1100											
Pro	Ser	Gln	Lys	Ala	Glu	Ile	Thr	Glu	Leu	Ser	Thr	Ile	Leu	Glu	Glu								
1105						1110						1115						1120					
Ser	Gly	Ser	Gln	Phe	Glu	Phe	Thr	Gln	Phe	Arg	Lys	Pro	Ser	Tyr	Ile								
						1125						1130						1135					
Leu	Gln	Lys	Ser	Thr	Phe	Glu	Val	Pro	Glu	Asn	Gln	Met	Thr	Ile	Leu								
1140						1145						1150											
Lys	Thr	Thr	Ser	Glu	Glu	Cys	Arg	Asp	Ala	Asp	Leu	His	Val	Ile	Met								
1155						1160						1165											
Asn	Ala	Pro	Ser	Ile	Gly	Gln	Val	Asp	Ser	Ser	Lys	Gln	Phe	Glu	Gly								
1170						1175						1180											
Thr	Val	Glu	Ile	Lys	Arg	Lys	Phe	Ala	Gly	Leu	Leu	Lys	Asn	Asp	Cys								
1185						1190						1195						1200					
Asn	Lys	Ser	Ala	Ser	Gly	Tyr	Leu	Thr	Asp	Glu	Asn	Glu	Val	Gly	Phe								
						1205						1210						1215					
Arg	Gly	Phe	Tyr	Ser	Ala	His	Gly	Thr	Lys	Leu	Asn	Val	Ser	Thr	Glu								
1220						1225						1230											
Ala	Leu	Gln	Lys	Ala	Val	Lys	Leu	Phe	Ser	Asp	Ile	Glu	Asn	Ile	Ser								
1235						1240						1245											
Glu	Glu	Thr	Ser	Ala	Glu	Val	His	Pro	Ile	Ser	Leu	Ser	Ser	Ser	Lys								
1250						1255						1260											

6,033,857

91

92

-continued

Cys His Asp Ser Val Val Ser Met Phe Lys Ile Glu Asn His Asn Asp
1265 1270 1275 1280

Lys Thr Val Ser Glu Lys Asn Asn Lys Cys Gln Leu Ile Leu Gln Asn
1285 1290 1295

Asn Ile Glu Met Thr Thr Gly Thr Phe Val Glu Glu Ile Thr Glu Asn
1300 1305 1310

Tyr Lys Arg Asn Thr Glu Asn Glu Asp Asn Lys Tyr Thr Ala Ala Ser
1315 1320 1325

Arg Asn Ser His Asn Leu Glu Phe Asp Gly Ser Asp Ser Ser Lys Asn
1330 1335 1340

Asp Thr Val Cys Ile His Lys Asp Glu Thr Asp Leu Leu Phe Thr Asp
1345 1350 1355 1360

Gln His Asn Ile Cys Leu Lys Leu Ser Gly Gln Phe Met Lys Glu Gly
1365 1370 1375

Asn Thr Gln Ile Lys Glu Asp Leu Ser Asp Leu Thr Phe Leu Glu Val
1380 1385 1390

Ala Lys Ala Gln Glu Ala Cys His Gly Asn Thr Ser Asn Lys Glu Gln
1395 1400 1405

Leu Thr Ala Thr Lys Thr Glu Gln Asn Ile Lys Asp Phe Glu Thr Ser
1410 1415 1420

Asp Thr Phe Phe Gln Thr Ala Ser Gly Lys Asn Ile Ser Val Ala Lys
1425 1430 1435 1440

Glu Ser Phe Asn Lys Ile Val Asn Phe Phe Asp Gln Lys Pro Glu Glu
1445 1450 1455

Leu His Asn Phe Ser Leu Asn Ser Glu Leu His Ser Asp Ile Arg Lys
1460 1465 1470

Asn Lys Met Asp Ile Leu Ser Tyr Glu Glu Thr Asp Ile Val Lys His
1475 1480 1485

Lys Ile Leu Lys Glu Ser Val Pro Val Gly Thr Gly Asn Gln Leu Val
1490 1495 1500

Thr Phe Gln Gly Gln Pro Glu Arg Asp Glu Lys Ile Lys Glu Pro Thr
1505 1510 1515 1520

Leu Leu Gly Phe His Thr Ala Ser Gly Lys Lys Val Lys Ile Ala Lys
1525 1530 1535

Glu Ser Leu Asp Lys Val Lys Asn Leu Phe Asp Glu Lys Glu Gln Gly
1540 1545 1550

Thr Ser Glu Ile Thr Ser Phe Ser His Gln Trp Ala Lys Thr Leu Lys
1555 1560 1565

Tyr Arg Glu Ala Cys Lys Asp Leu Glu Leu Ala Cys Glu Thr Ile Glu
1570 1575 1580

Ile Thr Ala Ala Pro Lys Cys Lys Glu Met Gln Asn Ser Leu Asn Asn
1585 1590 1595 1600

Asp Lys Asn Leu Val Ser Ile Glu Thr Val Val Pro Pro Lys Leu Leu
1605 1610 1615

Ser Asp Asn Leu Cys Arg Gln Thr Glu Asn Leu Lys Thr Ser Lys Ser
1620 1625 1630

Ile Phe Leu Lys Val Lys Val His Glu Asn Val Glu Lys Glu Thr Ala
1635 1640 1645

Lys Ser Pro Ala Thr Cys Tyr Thr Asn Gln Ser Pro Tyr Ser Val Ile
1650 1655 1660

Glu Asn Ser Ala Leu Ala Phe Tyr Thr Ser Cys Ser Arg Lys Thr Ser
1665 1670 1675 1680

-continued															
Val	Ser	Gln	Thr	Ser	Leu	Leu	Glu	Ala	Lys	Lys	Trp	Leu	Arg	Glu	Gly
				1685					1690					1695	
Ile	Phe	Asp	Gly	Gln	Pro	Glu	Arg	Ile	Asn	Thr	Ala	Asp	Tyr	Val	Gly
			1700					1705					1710		
Asn	Tyr	Leu	Tyr	Glu	Asn	Asn	Ser	Asn	Ser	Thr	Ile	Ala	Glu	Asn	Asp
		1715					1720					1725			
Lys	Asn	His	Leu	Ser	Glu	Lys	Gln	Asp	Thr	Tyr	Leu	Ser	Asn	Ser	Ser
		1730				1735					1740				
Met	Ser	Asn	Ser	Tyr	Ser	Tyr	His	Ser	Asp	Glu	Val	Tyr	Asn	Asp	Ser
				1750						1755					1760
Gly	Tyr	Leu	Ser	Lys	Asn	Lys	Leu	Asp	Ser	Gly	Ile	Glu	Pro	Val	Leu
				1765					1770					1775	
Lys	Asn	Val	Glu	Asp	Gln	Lys	Asn	Thr	Ser	Phe	Ser	Lys	Val	Ile	Ser
			1780					1785						1790	
Asn	Val	Lys	Asp	Ala	Asn	Ala	Tyr	Pro	Gln	Thr	Val	Asn	Glu	Asp	Ile
			1795					1800					1805		
Cys	Val	Glu	Glu	Leu	Val	Thr	Ser	Ser	Ser	Pro	Cys	Lys	Asn	Lys	Asn
		1810				1815					1820				
Ala	Ala	Ile	Lys	Leu	Ser	Ile	Ser	Asn	Ser	Asn	Asn	Phe	Glu	Val	Gly
		1825				1830				1835					1840
Pro	Pro	Ala	Phe	Arg	Ile	Ala	Ser	Gly	Lys	Ile	Val	Cys	Val	Ser	His
				1845					1850						1855
Glu	Thr	Ile	Lys	Lys	Val	Lys	Asp	Ile	Phe	Thr	Asp	Ser	Phe	Ser	Lys
			1860					1865						1870	
Val	Ile	Lys	Glu	Asn	Asn	Glu	Asn	Lys	Ser	Lys	Ile	Cys	Gln	Thr	Lys
		1875				1880						1885			
Ile	Met	Ala	Gly	Cys	Tyr	Glu	Ala	Leu	Asp	Asp	Ser	Glu	Asp	Ile	Leu
		1890				1895					1900				
His	Asn	Ser	Leu	Asp	Asn	Asp	Glu	Cys	Ser	Thr	His	Ser	His	Lys	Val
		1905				1910				1915					1920
Phe	Ala	Asp	Ile	Gln	Ser	Glu	Glu	Ile	Leu	Gln	His	Asn	Gln	Asn	Met
			1925						1930						1935
Ser	Gly	Leu	Glu	Lys	Val	Ser	Lys	Ile	Ser	Pro	Cys	Asp	Val	Ser	Leu
		1940						1945					1950		
Glu	Thr	Ser	Asp	Ile	Cys	Lys	Cys	Ser	Ile	Gly	Lys	Leu	His	Lys	Ser
		1955					1960						1965		
Val	Ser	Ser	Ala	Asn	Thr	Cys	Gly	Ile	Phe	Ser	Thr	Ala	Ser	Gly	Lys
		1970				1975						1980			
Ser	Val	Gln	Val	Ser	Asp	Ala	Ser	Leu	Gln	Asn	Ala	Arg	Gln	Val	Phe
		1985				1990				1995					2000
Ser	Glu	Ile	Glu	Asp	Ser	Thr	Lys	Gln	Val	Phe	Ser	Lys	Val	Leu	Phe
			2005					2010						2015	
Lys	Ser	Asn	Glu	His	Ser	Asp	Gln	Leu	Thr	Arg	Glu	Glu	Asn	Thr	Ala
			2020					2025						2030	
Ile	Arg	Thr	Pro	Glu	His	Leu	Ile	Ser	Gln	Lys	Gly	Phe	Ser	Tyr	Asn
		2035						2040					2045		
Val	Val	Asn	Ser	Ser	Ala	Phe	Ser	Gly	Phe	Ser	Thr	Ala	Ser	Gly	Lys
		2050				2055						2060			
Gln	Val	Ser	Ile	Leu	Glu	Ser	Ser	Leu	His	Lys	Val	Lys	Gly	Val	Leu
		2065				2070				2075					2080
Glu	Glu	Phe	Asp	Leu	Ile	Arg	Thr	Glu	His	Ser	Leu	His	Tyr	Ser	Pro
			2085						2090					2095	
Thr	Ser	Arg	Gln	Asn	Val	Ser	Lys	Ile	Leu	Pro	Arg	Val	Asp	Lys	Arg

6,033,857

95

96

-continued

2100					2105					2110						
Asn	Pro	Glu	His	Cys	Val	Asn	Ser	Glu	Met	Glu	Lys	Thr	Cys	Ser	Lys	
2115					2120					2125						
Glu	Phe	Lys	Leu	Ser	Asn	Asn	Leu	Asn	Val	Glu	Gly	Gly	Ser	Ser	Glu	
2130					2135					2140						
Asn	Asn	His	Ser	Ile	Lys	Val	Ser	Pro	Tyr	Leu	Ser	Gln	Phe	Gln	Gln	
2145					2150					2155					2160	
Asp	Lys	Gln	Gln	Leu	Val	Leu	Gly	Thr	Lys	Val	Ser	Leu	Val	Glu	Asn	
2165					2170					2175						
Ile	His	Val	Leu	Gly	Lys	Glu	Gln	Ala	Ser	Pro	Lys	Asn	Val	Lys	Met	
2180					2185					2190						
Glu	Ile	Gly	Lys	Thr	Glu	Thr	Phe	Ser	Asp	Val	Pro	Val	Lys	Thr	Asn	
2195					2200					2205						
Ile	Glu	Val	Cys	Ser	Thr	Tyr	Ser	Lys	Asp	Ser	Glu	Asn	Tyr	Phe	Glu	
2210					2215					2220						
Thr	Glu	Ala	Val	Glu	Ile	Ala	Lys	Ala	Phe	Met	Glu	Asp	Asp	Glu	Leu	
2225					2230					2235					2240	
Thr	Asp	Ser	Lys	Leu	Pro	Ser	His	Ala	Thr	His	Ser	Leu	Phe	Thr	Cys	
2245					2250					2255						
Pro	Glu	Asn	Glu	Glu	Met	Val	Leu	Ser	Asn	Ser	Arg	Ile	Gly	Lys	Arg	
2260					2265					2270						
Arg	Gly	Glu	Pro	Leu	Ile	Leu	Val	Gly	Glu	Pro	Ser	Ile	Lys	Arg	Asn	
2275					2280					2285						
Leu	Leu	Asn	Glu	Phe	Asp	Arg	Ile	Ile	Glu	Asn	Gln	Glu	Lys	Ser	Leu	
2290					2295					2300						
Lys	Ala	Ser	Lys	Ser	Thr	Pro	Asp	Gly	Thr	Ile	Lys	Asp	Arg	Arg	Leu	
2305					2310					2315					2320	
Phe	Met	His	His	Val	Ser	Leu	Glu	Pro	Ile	Thr	Cys	Val	Pro	Phe	Arg	
2325					2330					2335						
Thr	Thr	Lys	Glu	Arg	Gln	Glu	Ile	Gln	Asn	Pro	Asn	Phe	Thr	Ala	Pro	
2340					2345					2350						
Gly	Gln	Glu	Phe	Leu	Ser	Lys	Ser	His	Leu	Tyr	Glu	His	Leu	Thr	Leu	
2355					2360					2365						
Glu	Lys	Ser	Ser	Ser	Asn	Leu	Ala	Val	Ser	Gly	His	Pro	Phe	Tyr	Gln	
2370					2375					2380						
Val	Ser	Ala	Thr	Arg	Asn	Glu	Lys	Met	Arg	His	Leu	Ile	Thr	Thr	Gly	
2385					2390					2395					2400	
Arg	Pro	Thr	Lys	Val	Phe	Val	Pro	Pro	Phe	Lys	Thr	Lys	Ser	His	Phe	
2405					2410					2415						
His	Arg	Val	Glu	Gln	Cys	Val	Arg	Asn	Ile	Asn	Leu	Glu	Glu	Asn	Arg	
2420					2425					2430						
Gln	Lys	Gln	Asn	Ile	Asp	Gly	His	Gly	Ser	Asp	Asp	Ser	Lys	Asn	Lys	
2435					2440					2445						
Ile	Asn	Asp	Asn	Glu	Ile	His	Gln	Phe	Asn	Lys	Asn	Asn	Ser	Asn	Gln	
2450					2455					2460						
Ala	Ala	Ala	Val	Thr	Phe	Thr	Lys	Cys	Glu	Glu	Glu	Pro	Leu	Asp	Leu	
2465					2470					2475					2480	
Ile	Thr	Ser	Leu	Gln	Asn	Ala	Arg	Asp	Ile	Gln	Asp	Met	Arg	Ile	Lys	
2485					2490					2495						
Lys	Lys	Gln	Arg	Gln	Arg	Val	Phe	Pro	Gln	Pro	Gly	Ser	Leu	Tyr	Leu	
2500					2505					2510						
Ala	Lys	Thr	Ser	Thr	Leu	Pro	Arg	Ile	Ser	Leu	Lys	Ala	Ala	Val	Gly	
2515					2520					2525						

6,033,857

97

98

-continued

Gly Gln Val Pro Ser Ala Cys Ser His Lys Gln Leu Tyr Thr Tyr Gly
 2530 2535 2540
 Val Ser Lys His Cys Ile Lys Ile Asn Ser Lys Asn Ala Glu Ser Phe
 2545 2550 2555 2560
 Gln Phe His Thr Glu Asp Tyr Phe Gly Lys Glu Ser Leu Trp Thr Gly
 2565 2570 2575
 Lys Gly Ile Gln Leu Ala Asp Gly Gly Trp Leu Ile Pro Ser Asn Asp
 2580 2585 2590
 Gly Lys Ala Gly Lys Glu Glu Phe Tyr Arg Ala Leu Cys Asp Thr Pro
 2595 2600 2605
 Gly Val Asp Pro Lys Leu Ile Ser Arg Ile Trp Val Tyr Asn His Tyr
 2610 2615 2620
 Arg Trp Ile Ile Trp Lys Leu Ala Ala Met Glu Cys Ala Phe Pro Lys
 2625 2630 2635 2640
 Glu Phe Ala Asn Arg Cys Leu Ser Pro Glu Arg Val Leu Leu Gln Leu
 2645 2650 2655
 Lys Tyr Arg Tyr Asp Thr Glu Ile Asp Arg Ser Arg Arg Ser Ala Ile
 2660 2665 2670
 Lys Lys Ile Met Glu Arg Asp Asp Thr Ala Ala Lys Thr Leu Val Leu
 2675 2680 2685
 Cys Val Ser Asp Ile Ile Ser Leu Ser Ala Asn Ile Ser Glu Thr Ser
 2690 2695 2700
 Ser Asn Lys Thr Ser Ser Ala Asp Thr Gln Lys Val Ala Ile Ile Glu
 2705 2710 2715 2720
 Leu Thr Asp Gly Trp Tyr Ala Val Lys Ala Gln Leu Asp Pro Pro Leu
 2725 2730 2735
 Leu Ala Val Leu Lys Asn Gly Arg Leu Thr Val Gly Gln Lys Ile Ile
 2740 2745 2750
 Leu His Gly Ala Glu Leu Val Gly Ser Pro Asp Ala Cys Thr Pro Leu
 2755 2760 2765
 Glu Ala Pro Glu Ser Leu Met Leu Lys Ile Ser Ala Asn Ser Thr Arg
 2770 2775 2780
 Pro Ala Arg Trp Tyr Thr Lys Leu Gly Phe Phe Pro Asp Pro Arg Pro
 2785 2790 2795 2800
 Phe Pro Leu Pro Leu Ser Ser Leu Phe Ser Asp Gly Gly Asn Val Gly
 2805 2810 2815
 Cys Val Asp Val Ile Ile Gln Arg Ala Tyr Pro Ile Gln Trp Met Glu
 2820 2825 2830
 Lys Thr Ser Ser Gly Leu Tyr Ile Phe Arg Asn Glu Arg Glu Glu Glu
 2835 2840 2845
 Lys Glu Ala Ala Lys Tyr Val Glu Ala Gln Gln Lys Arg Leu Glu Ala
 2850 2855 2860
 Leu Phe Thr Lys Ile Gln Glu Glu Phe Glu Glu His Glu Glu Asn Thr
 2865 2870 2875 2880
 Thr Lys Pro Tyr Leu Pro Ser Arg Ala Leu Thr Arg Gln Gln Val Arg
 2885 2890 2895
 Ala Leu Gln Asp Gly Ala Glu Leu Tyr Glu Ala Val Lys Asn Ala Ala
 2900 2905 2910
 Asp Pro Ala Tyr Leu Glu Gly Tyr Phe Ser Glu Glu Gln Leu Arg Ala
 2915 2920 2925
 Leu Asn Asn His Arg Gln Met Leu Asn Asp Lys Lys Gln Ala Gln Ile
 2930 2935 2940

A000506

6,033,857

99

100

-continued

Gln	Leu	Glu	Ile	Arg	Lys	Ala	Met	Glu	Ser	Ala	Glu	Gln	Lys	Glu	Gln
2945					2950					2955					2960
Gly	Leu	Ser	Arg	Asp	Val	Thr	Thr	Val	Trp	Lys	Leu	Arg	Ile	Val	Ser
			2965						2970					2975	
Tyr	Ser	Lys	Lys	Glu	Lys	Asp	Ser	Val	Ile	Leu	Ser	Ile	Trp	Arg	Pro
			2980					2985						2990	
Ser	Ser	Asp	Leu	Tyr	Ser	Leu	Leu	Thr	Glu	Gly	Lys	Arg	Tyr	Arg	Ile
		2995					3000					3005			
Tyr	His	Leu	Ala	Thr	Ser	Lys	Ser	Lys	Ser	Lys	Ser	Glu	Arg	Ala	Asn
	3010					3015						3020			
Ile	Gln	Leu	Ala	Ala	Thr	Lys	Lys	Thr	Gln	Tyr	Gln	Gln	Leu	Pro	Val
3025					3030					3035					3040
Ser	Asp	Glu	Ile	Leu	Phe	Gln	Ile	Tyr	Gln	Pro	Arg	Glu	Pro	Leu	His
				3045					3050					3055	
Phe	Ser	Lys	Phe	Leu	Asp	Pro	Asp	Phe	Gln	Pro	Ser	Cys	Ser	Glu	Val
			3060					3065					3070		
Asp	Leu	Ile	Gly	Phe	Val	Val	Ser	Val	Val	Lys	Lys	Thr	Gly	Leu	Ala
	3075						3080					3085			
Pro	Phe	Val	Tyr	Leu	Ser	Asp	Glu	Cys	Tyr	Asn	Leu	Leu	Ala	Ile	Lys
	3090					3095					3100				
Phe	Trp	Ile	Asp	Leu	Asn	Glu	Asp	Ile	Ile	Lys	Pro	His	Met	Leu	Ile
3105					3110					3115					3120
Ala	Ala	Ser	Asn	Leu	Gln	Trp	Arg	Pro	Glu	Ser	Lys	Ser	Gly	Leu	Leu
			3125						3130					3135	
Thr	Leu	Phe	Ala	Gly	Asp	Phe	Ser	Val	Phe	Ser	Ala	Ser	Pro	Lys	Glu
	3140						3145						3150		
Gly	His	Phe	Gln	Glu	Thr	Phe	Asn	Lys	Met	Lys	Asn	Thr	Val	Glu	Asn
	3155						3160					3165			
Ile	Asp	Ile	Leu	Cys	Asn	Glu	Ala	Glu	Asn	Lys	Leu	Met	His	Ile	Leu
	3170				3175						3180				
His	Ala	Asn	Asp	Pro	Lys	Trp	Ser	Thr	Pro	Thr	Lys	Asp	Cys	Thr	Ser
3185					3190					3195					3200
Gly	Pro	Tyr	Thr	Ala	Gln	Ile	Ile	Pro	Gly	Thr	Gly	Asn	Lys	Leu	Leu
			3205						3210					3215	
Met	Ser	Ser	Pro	Asn	Cys	Glu	Ile	Tyr	Tyr	Gln	Ser	Pro	Leu	Ser	Leu
			3220					3225					3230		
Cys	Met	Ala	Lys	Arg	Lys	Ser	Val	Ser	Thr	Pro	Val	Ser	Ala	Gln	Met
	3235						3240					3245			
Thr	Ser	Lys	Ser	Cys	Lys	Gly	Glu	Lys	Glu	Ile	Asp	Asp	Gln	Lys	Asn
	3250				3255						3260				
Cys	Lys	Lys	Arg	Arg	Ala	Leu	Asp	Phe	Leu	Ser	Arg	Leu	Pro	Leu	Pro
3265					3270					3275					3280
Pro	Pro	Val	Ser	Pro	Ile	Cys	Thr	Phe	Val	Ser	Pro	Ala	Ala	Gln	Lys
			3285					3290						3295	
Ala	Phe	Gln	Pro	Pro	Arg	Ser	Cys	Gly	Thr	Lys	Tyr	Glu	Thr	Pro	Ile
		3300						3305					3310		
Lys	Lys	Lys	Glu	Leu	Asn	Ser	Pro	Gln	Met	Thr	Pro	Phe	Lys	Lys	Phe
	3315					3320						3325			
Asn	Glu	Ile	Ser	Leu	Leu	Glu	Ser	Asn	Ser	Ile	Ala	Asp	Glu	Glu	Leu
	3330				3335						3340				
Ala	Leu	Ile	Asn	Thr	Gln	Ala	Leu	Leu	Ser	Gly	Ser	Thr	Gly	Glu	Lys
3345					3350					3355					3360
Gln	Phe	Ile	Ser	Val	Ser	Glu	Ser	Thr	Arg	Thr	Ala	Pro	Thr	Ser	Ser

6,033,857

101

102

-continued

3365										3370										3375									
Glu	Asp	Tyr	Leu	Arg	Leu	Lys	Arg	Arg	Cys	Thr	Thr	Ser	Leu	Ile	Lys														
			3380						3385					3390															
Glu	Gln	Glu	Ser	Ser	Gln	Ala	Ser	Thr	Glu	Glu	Cys	Glu	Lys	Asn	Lys														
			3395					3400					3405																
Gln	Asp	Thr	Ile	Thr	Thr	Lys	Lys	Tyr	Ile																				
	3410						3415																						

- (2) INFORMATION FOR SEQ ID NO:3:
- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 32 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: DNA (genomic)
 - (iii) HYPOTHETICAL: NO
 - (iv) ANTI-SENSE: NO
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo sapiens
 - (ix) FEATURE:
 - (A) NAME/KEY: misc_feature
 - (B) LOCATION: 1..2
 - (D) OTHER INFORMATION: /note= "(NH2) at nucleotide 1"
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

GTAGTGCAAG GCTCGAGAAC NNNNNNNNNN NN 32

- (2) INFORMATION FOR SEQ ID NO:4:
- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 30 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: other nucleic acid
 - (A) DESCRIPTION: /desc = "primer"
 - (iii) HYPOTHETICAL: NO
 - (iv) ANTI-SENSE: NO
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo sapiens
 - (ix) FEATURE:
 - (A) NAME/KEY: misc_feature
 - (B) LOCATION: 1..2
 - (D) OTHER INFORMATION: /note= "(NH2) at nucleotide 1"
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

TGAGTAGAAT TCTAACGGCC GTCATTGTTC 30

- (2) INFORMATION FOR SEQ ID NO:5:
- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 30 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: other nucleic acid
 - (A) DESCRIPTION: /desc = "primer"
 - (iii) HYPOTHETICAL: NO

-continued

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens

(ix) FEATURE:

(A) NAME/KEY: misc_feature

(B) LOCATION: 29..30

(D) OTHER INFORMATION: /note= "(NH2) at nucleotide 30"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

GAACAATGAC GGCCGTTAGA ATTCTACTCA

30

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "primer"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

TCAGTAGAAT TCTAACGGCC GTCAT

25

INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "primer"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens

(ix) FEATURE:

(A) NAME/KEY: misc_feature

(B) LOCATION: 1..2

(D) OTHER INFORMATION: /note= "(PO4) at nucleotide 1"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

GTAGTGCAAG GCTCGAGAAC

20

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 27 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "primer"

(iii) HYPOTHETICAL: NO

6,033,857

105

106

-continued

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 1..2
(D) OTHER INFORMATION: /note= "(PO4) at nucleotide 1"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

TGAGTAGAAT TCTAACGGCC GTCATTG 27

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 33 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "primer"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 32..33
(D) OTHER INFORMATION: /note= "(NH2) at nucleotide 33"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

CCTTCACACG CGTATCGATT AGTCACNNNN NNN 33

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "primer"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 1..2
(D) OTHER INFORMATION: /note= "(PO4) at nucleotide 1"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

GTGACTAATC GATACGCGTG TGAAGGTGC 29

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

6,033,857

107

108

-continued

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "primer"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homos sapiens

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 1..2
(D) OTHER INFORMATION: /note= "Biotinylated at nucleotide 1"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

TTGAAGAACA ACAGGACTTT CACTA 25

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "primer"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

CACCTTCACA CGCGTATCG 19

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "primer"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

GTTTCGTAATT GTTGTTTTTA TGTTCAG 27

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "primer"

6,033,857

109

110

-continued

<hr/>	
(iii) HYPOTHETICAL: NO	
(iv) ANTI-SENSE: NO	
(vi) ORIGINAL SOURCE:	
(A) ORGANISM: Homo sapiens	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:	
CCTTCACACG CGTATCGATT AG	22
(2) INFORMATION FOR SEQ ID NO:15:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 22 base pairs	
(B) TYPE: nucleic acid	
(C) STRANDEDNESS: single	
(D) TOPOLOGY: linear	
(ii) MOLECULE TYPE: other nucleic acid	
(A) DESCRIPTION: /desc = "primer"	
(iii) HYPOTHETICAL: NO	
(iv) ANTI-SENSE: NO	
(vi) ORIGINAL SOURCE:	
(A) ORGANISM: Homo sapiens	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:	
TTTGGATCAT TTTACACTG TC	22
(2) INFORMATION FOR SEQ ID NO:16:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 22 base pairs	
(B) TYPE: nucleic acid	
(C) STRANDEDNESS: single	
(D) TOPOLOGY: linear	
(ii) MOLECULE TYPE: other nucleic acid	
(A) DESCRIPTION: /desc = "primer"	
(iii) HYPOTHETICAL: NO	
(iv) ANTI-SENSE: NO	
(vi) ORIGINAL SOURCE:	
(A) ORGANISM: Homo sapiens	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:	
GTGCTCATAG TCAGAAATGA AG	22
(2) INFORMATION FOR SEQ ID NO:17:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 21 base pairs	
(B) TYPE: nucleic acid	
(C) STRANDEDNESS: single	
(D) TOPOLOGY: linear	
(ii) MOLECULE TYPE: DNA (genomic)	
(iii) HYPOTHETICAL: NO	
(iv) ANTI-SENSE: NO	
(vi) ORIGINAL SOURCE:	
(A) ORGANISM: Homo sapiens	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:	
TCTTCCCATC CTCACAGTAA G	21

6,033,857

111

112

-continued

(2) INFORMATION FOR SEO ID NO:18:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

GTACTGGGTT TTTAGCAAGC A

21

(2) INFORMATION FOR SEO ID NO:19:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEO ID NO:19:

GGTTAAACT AAGGTGGGA

19

(2) INFORMATION FOR SEO ID NO:20:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

ATTTGCCAG CATGACACA

19

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

-continued

(iv) ANTI-SENSE: NO	
(vi) ORIGINAL SOURCE:	
(A) ORGANISM: Homo sapiens	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:	
TTTCCCGTA TAGAGGAGA	19
(2) INFORMATION FOR SEQ ID NO:22:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 21 base pairs	
(B) TYPE: nucleic acid	
(C) STRANDEDNESS: single	
(D) TOPOLOGY: linear	
(ii) MOLECULE TYPE: DNA (genomic)	
(iii) HYPOTHETICAL: NO	
(iv) ANTI-SENSE: NO	
(vi) ORIGINAL SOURCE:	
(A) ORGANISM: Homo sapiens	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:	
GTAGGAAAAT GTTTCATTA A	21
(2) INFORMATION FOR SEQ ID NO:23:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 21 base pairs	
(B) TYPE: nucleic acid	
(C) STRANDEDNESS: single	
(D) TOPOLOGY: linear	
(ii) MOLECULE TYPE: DNA (genomic)	
(iii) HYPOTHETICAL: NO	
(iv) ANTI-SENSE: NO	
(vi) ORIGINAL SOURCE:	
(A) ORGANISM: Homo sapiens	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:	
ATCTAAAGTA GTATTCCAAC A	21
(2) INFORMATION FOR SEQ ID NO:24:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 19 base pairs	
(B) TYPE: nucleic acid	
(C) STRANDEDNESS: single	
(D) TOPOLOGY: linear	
(ii) MOLECULE TYPE: DNA (genomic)	
(iii) HYPOTHETICAL: NO	
(iv) ANTI-SENSE: NO	
(vi) ORIGINAL SOURCE:	
(A) ORGANISM: Homo sapiens	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:	
GGGGGTAAAA AAAGGGGAA	19
(2) INFORMATION FOR SEQ ID NO:25:	
(i) SEQUENCE CHARACTERISTICS:	

6,033,857

115

116

-continued

(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

GAGATAAGTC AGGTATGATT 20

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

AATTGCCTGT ATGAGGCAGA 20

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

GGCAATTCAG TAAACGTAA 20

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

6,033,857

117

118

-continued

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:
ATTGTCAGTT ACTAACACAC 20

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:
GTGTCATGTA ATCAAATAGT 20

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:
CAGGTTTAGA GACTTTCTC 19

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:
GGACCTAGGT TGATTGCA 18

(2) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

6,033,857

119

120

-continued

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

GTCAAGAAAG GTAAGGTAA 19

(2) INFORMATION FOR SEQ ID NO:33:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

CTATGAGAAA GGTGTGTAG 19

(2) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

CCTAGTCTTG CTAGTTCTT 19

(2) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

AACAGTTGTA GATACCTCTG AA 22

6,033,857

121

122

-continued

(2) INFORMATION FOR SEO ID NO:36:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 22 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEO ID NO:36:

GACTTTTGA TACCCTGAAA TG

22

(2) INFORMATION FOR SEQ ID NO:37:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

CAGCATCTTG AATCTCATAC AG

22

(2) INFORMATION FOR SEQ ID NO:38:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 23 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

CATGTATACA GATGATGCCT AAG

23

(2) INFORMATION FOR SEQ ID NO:39:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

-continued

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

AACTTAGTGA AAAATATTTA GTGA24

(2) INFORMATION FOR SEQ ID NO:40:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

ATACATCTTG ATTCTTTTCC AT22

(2) INFORMATION FOR SEQ ID NO:41:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

TTTAGTGAAT GTGATTGATG GT22

(2) INFORMATION FOR SEQ ID NO:42:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

AGAACCAACT TTGTCCTTAA20

(2) INFORMATION FOR SEQ ID NO:43:

6,033,857

125

126

-continued

<div>(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 22 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear</div> <div>(ii) MOLECULE TYPE: DNA (genomic)</div> <div>(iii) HYPOTHETICAL: NO</div> <div>(iv) ANTI-SENSE: NO</div> <div>(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens</div> <div>(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:</div> <div>TTAGATTGTGTTTGGTTG AA</div>		22
<div>(2) INFORMATION FOR SEQ ID NO:44:</div> <div>(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 20 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear</div> <div>(ii) MOLECULE TYPE: DNA (genomic)</div> <div>(iii) HYPOTHETICAL: NO</div> <div>(iv) ANTI-SENSE: NO</div> <div>(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens</div> <div>(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:</div> <div>TAGCTCTTT GGGACAATTC</div>		20
<div>(2) INFORMATION FOR SEQ ID NO:45:</div> <div>(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 22 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear</div> <div>(ii) MOLECULE TYPE: DNA (genomic)</div> <div>(iii) HYPOTHETICAL: NO</div> <div>(iv) ANTI-SENSE: NO</div> <div>(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens</div> <div>(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:</div> <div>ATGGAAAAGA ATCAAGATGT AT</div>		22
<div>(2) INFORMATION FOR SEQ ID NO:46:</div> <div>(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 21 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear</div> <div>(ii) MOLECULE TYPE: DNA (genomic)</div> <div>(iii) HYPOTHETICAL: NO</div> <div>(iv) ANTI-SENSE: NO</div> <div>(vi) ORIGINAL SOURCE:</div>		

6,033,857

127

128

-continued

(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

CCTAATGTTA TGTCAGAGA G 21

(2) INFORMATION FOR SEQ ID NO:47:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 19 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:

GCTACCTCCA AAACGTGA 19

(2) INFORMATION FOR SEQ ID NO:48:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 22 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:

GTGTAAAGCA GCATATAAAA AT 22

(2) INFORMATION FOR SEQ ID NO:49:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 18 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

CTTGCTGCTG TCTACCTG 18

(2) INFORMATION FOR SEQ ID NO:50:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

A000521

6,033,857

129

130

-continued

```

(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: DNA (genomic)
(iii) HYPOTHETICAL: NO
(iv) ANTI-SENSE: NO
(vi) ORIGINAL SOURCE:
      (A) ORGANISM: Homo sapiens
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:

AGTGGTCTTA AGATAGTCAT
20

(2) INFORMATION FOR SEQ ID NO:51:
(i) SEQUENCE CHARACTERISTICS:
      (A) LENGTH: 21 base pairs
      (B) TYPE: nucleic acid
      (C) STRANDEDNESS: single
      (D) TOPOLOGY: linear
(ii) MOLECULE TYPE: DNA (genomic)
(iii) HYPOTHETICAL: NO
(iv) ANTI-SENSE: NO
(vi) ORIGINAL SOURCE:
      (A) ORGANISM: Homo sapiens
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:

CCATAATTTA ACACCTAGCC A
21

(2) INFORMATION FOR SEQ ID NO:52:
(i) SEQUENCE CHARACTERISTICS:
      (A) LENGTH: 21 base pairs
      (B) TYPE: nucleic acid
      (C) STRANDEDNESS: single
      (D) TOPOLOGY: linear
(ii) MOLECULE TYPE: DNA (genomic)
(iii) HYPOTHETICAL: NO
(iv) ANTI-SENSE: NO
(vi) ORIGINAL SOURCE:
      (A) ORGANISM: Homo sapiens
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:52:

CCAAAAAAGT TAAATCTGAC A
21

(2) INFORMATION FOR SEQ ID NO:53:
(i) SEQUENCE CHARACTERISTICS:
      (A) LENGTH: 21 base pairs
      (B) TYPE: nucleic acid
      (C) STRANDEDNESS: single
      (D) TOPOLOGY: linear
(ii) MOLECULE TYPE: DNA (genomic)
(iii) HYPOTHETICAL: NO
(iv) ANTI-SENSE: NO
(vi) ORIGINAL SOURCE:
      (A) ORGANISM: Homo sapiens
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:53:

```

6,033,857

131

132

-continued

GGCTTTTATT CTGCTCATGG C

21

(2) INFORMATION FOR SEO ID NO:54:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:54:

CCTCTGCAGA AGTTTCCTCA C

21

(2) INFORMATION FOR SEQ ID NO:55:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:55:

AACGGACTTG CTATTTACTG A

21

(2) INFORMATION FOR SEQ ID NO:56:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEO ID NO:56:

AGTACCTTGC TCTTTTTCAT C

21

(2) INFORMATION FOR SEQ ID NO:57:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

6,033,857

133

134

-continued

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:57:

CAGCTAGCGG GAAAAAAGTT A

(2) INFORMATION FOR SEQ ID NO:58:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 21 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo sapiens
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:58:

TTCGGAGAGA TGATTTTGT C

(2) INFORMATION FOR SEQ ID NO:59:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 19 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo sapiens
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:59:

GCCTTAGCTT TTTACACA

```
(2) INFORMATION FOR SEQ ID NO:60:

      (i) SEQUENCE CHARACTERISTICS:
            (A) LENGTH: 20 base pairs
            (B) TYPE: nucleic acid
            (C) STRANDEDNESS: single
            (D) TOPOLOGY: linear

      (ii) MOLECULE TYPE: DNA (genomic)

      (iii) HYPOTHETICAL: NO

      (iv) ANTI-SENSE: NO

      (vi) ORIGINAL SOURCE:
            (A) ORGANISM: Homo sapiens

      (xi) SEQUENCE DESCRIPTION: SEQ ID NO:60:

TTTTTGATTA TATCTCGTTG
```

(2) INFORMATION FOR SEQ ID NO:61:

6,033,857

135

136

-continued

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 21 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
 (A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:61:

TTATTCTCGT TGTTTTCTTT A 21

(2) INFORMATION FOR SEQ ID NO:62:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 21 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
 (A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:62:

CCATTAAATT GTCCATATCT A 21

(2) INFORMATION FOR SEQ ID NO:63:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 21 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
 (A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:63:

GACGTAGGTG AATAGTGAAG A 21

(2) INFORMATION FOR SEQ ID NO:64:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 20 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

6,033,857

137

138

-continued

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:64:

TCAAATTCCT CTAACACTCC 20

(2) INFORMATION FOR SEQ ID NO:65:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:65:

GAAGATAGTA CCAAGCAAGT C 21

(2) INFORMATION FOR SEQ ID NO:66:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:66:

TGAGACTTTG GTTCCTAATA C 21

(2) INFORMATION FOR SEQ ID NO:67:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:67:

AGTAACGAAC ATTCAGACCA G 21

(2) INFORMATION FOR SEQ ID NO:68:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid

A000526

6,033,857

139

140

-continued

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:68:

GTCTTCACTA TTCACCTACG 20

(2) INFORMATION FOR SEQ ID NO:69:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:69:

CCCCCAAAC TACTACACA 20

(2) INFORMATION FOR SEQ ID NO:70:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:70:

AGCATACCAA GTCTACTGAA T 21

(2) INFORMATION FOR SEQ ID NO:71:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:71:

A000527

6,033,857

141

142

-continued

ACTCTTTCAA ACATTAGGTC A

21

(2) INFORMATION FOR SEO ID NO:72:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEO ID NO:72:

TTGGAGAGGC AGGTGGAT

18

(2) INFORMATION FOR SEQ ID NO:73:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:73:

CTATAGAGGG AGAACAGAT

19

(2) INFORMATION FOR SEQ ID NO:74:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEO ID NO:74:

TTTATGCTGA TTTCTGTTGT AT

22

(2) INFORMATION FOR SEQ ID NO:75:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 21 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

6,033,857

143

144

-continued

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:75:

ATAAAACGGG AAGTGTTAAC T 21

(2) INFORMATION FOR SEQ ID NO:76:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:76:

CTGTGAGTGA TTTGGTGCAAT 20

(2) INFORMATION FOR SEQ ID NO:77:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:77:

GAATACAAAA CAGTACCAG A 21

(2) INFORMATION FOR SEQ ID NO:78:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:78:

CACCACCAAA GGGGGAAA 18

6,033,857

145

146

-continued

(2) INFORMATION FOR SEQ ID NO:79:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 20 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: *Homo sapiens*

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:79:

AAATGAGGGT CTGCAACAAA

20

(2) INFORMATION FOR SEO ID NO:80:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEO ID NO:80:

GTCCGACCAG AACTTGAG

18

(2) INFORMATION FOR SEQ ID NO:81:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:81:

AGCCATTGT AGGATACTAG

20

(2) INFORMATION FOR SEQ ID NO:82:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

6,033,857

147

148

-continued

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:82:

CTACTAGACG GCGCGAG 17

(2) INFORMATION FOR SEQ ID NO:83:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:83:

ATGTTTTTGT AGTGAAGATT CT 22

(2) INFORMATION FOR SEQ ID NO:84:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:84:

TAGTTCGAGA GACAGTTAAG 20

(2) INFORMATION FOR SEQ ID NO:85:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:85:

CAGTTTTTGT TTGTTATAAT TG 22

(2) INFORMATION FOR SEQ ID NO:86:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs

6,033,857

149

150

-continued

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:86:

CAGAGAATAG TTGTAGTTGT T 21

(2) INFORMATION FOR SEQ ID NO:87:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:87:

AACCTTAACC CATACTGCC 19

(2) INFORMATION FOR SEQ ID NO:88:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:88:

TTCAGTATCA TCCTATGTGG 20

(2) INFORMATION FOR SEQ ID NO:89:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

A000532

6,033,857

151

152

-continued

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:89:	
TTTTATTCTC AGTTATTCAG TG	22
(2) INFORMATION FOR SEQ ID NO:90:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 21 base pairs	
(B) TYPE: nucleic acid	
(C) STRANDEDNESS: single	
(D) TOPOLOGY: linear	
(ii) MOLECULE TYPE: DNA (genomic)	
(iii) HYPOTHETICAL: NO	
(iv) ANTI-SENSE: NO	
(vi) ORIGINAL SOURCE:	
(A) ORGANISM: Homo sapiens	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:90:	
GAAATTGAGC ATCCTTAGTA A	21
(2) INFORMATION FOR SEQ ID NO:91:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 20 base pairs	
(B) TYPE: nucleic acid	
(C) STRANDEDNESS: single	
(D) TOPOLOGY: linear	
(ii) MOLECULE TYPE: DNA (genomic)	
(iii) HYPOTHETICAL: NO	
(iv) ANTI-SENSE: NO	
(vi) ORIGINAL SOURCE:	
(A) ORGANISM: Homo sapiens	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:91:	
AATTCTAGAG TCACACTTCC	20
(2) INFORMATION FOR SEQ ID NO:92:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 22 base pairs	
(B) TYPE: nucleic acid	
(C) STRANDEDNESS: single	
(D) TOPOLOGY: linear	
(ii) MOLECULE TYPE: DNA (genomic)	
(iii) HYPOTHETICAL: NO	
(iv) ANTI-SENSE: NO	
(vi) ORIGINAL SOURCE:	
(A) ORGANISM: Homo sapiens	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:92:	
ATATTTTIAA GGCAGTTCTA GA	22
(2) INFORMATION FOR SEQ ID NO:93:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 21 base pairs	
(B) TYPE: nucleic acid	
(C) STRANDEDNESS: single	
(D) TOPOLOGY: linear	

6,033,857

153

154

-continued

```
(ii) MOLECULE TYPE: DNA (genomic)
(iii) HYPOTHETICAL: NO
(iv) ANTI-SENSE: NO
(vi) ORIGINAL SOURCE:
      (A) ORGANISM: Homo sapiens
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:93:

TTACACACAC CAAAAAAGTC A
```

(2) INFORMATION FOR SEQ ID NO:94:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 22 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo sapiens
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:94:

TGAAAACTCT TATGATATCT GT

(2) INFORMATION FOR SEQ ID NO:95:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 23 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo sapiens
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:95:

TGAATGTTAT ATATGTGACT TTT

(2) INFORMATION FOR SEQ ID NO:96:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 21 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo sapiens
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:96:

CTTGTTGCTA TTCTTGCT A

21

6,033,857

155

156

-continued

(2) INFORMATION FOR SEO ID NO:97:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:97:

CCCTAGATAC TAAAAAATAA AG

22

(2) INFORMATION FOR SEO ID NO:98:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEO ID NO:98:

CTTTTAGCAG TTATATAGTT TC

22

(2) INFORMATION FOR SEO ID NO:99:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:99:

GCCAGAGAGT CTAAAACAG

19

(2) INFORMATION FOR SEQ ID NO:100:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

6,033,857

157

158

-continued

(iv) ANTI-SENSE: NO	
(vi) ORIGINAL SOURCE:	
(A) ORGANISM: Homo sapiens	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:100:	
CTTTGGGTGT TTTATGCTTG	20
(2) INFORMATION FOR SEQ ID NO:101:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 22 base pairs	
(B) TYPE: nucleic acid	
(C) STRANDEDNESS: single	
(D) TOPOLOGY: linear	
(ii) MOLECULE TYPE: DNA (genomic)	
(iii) HYPOTHETICAL: NO	
(iv) ANTI-SENSE: NO	
(vi) ORIGINAL SOURCE:	
(A) ORGANISM: Homo sapiens	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:101:	
TTTGTGTAT TTGTCCTGTT TA	22
(2) INFORMATION FOR SEQ ID NO:102:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 23 base pairs	
(B) TYPE: nucleic acid	
(C) STRANDEDNESS: single	
(D) TOPOLOGY: linear	
(ii) MOLECULE TYPE: DNA (genomic)	
(iii) HYPOTHETICAL: NO	
(iv) ANTI-SENSE: NO	
(vi) ORIGINAL SOURCE:	
(A) ORGANISM: Homo sapiens	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:102:	
ATTTTGTTAG TAAGTCATT TTT	23
(2) INFORMATION FOR SEQ ID NO:103:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 21 base pairs	
(B) TYPE: nucleic acid	
(C) STRANDEDNESS: single	
(D) TOPOLOGY: linear	
(ii) MOLECULE TYPE: DNA (genomic)	
(iii) HYPOTHETICAL: NO	
(iv) ANTI-SENSE: NO	
(vi) ORIGINAL SOURCE:	
(A) ORGANISM: Homo sapiens	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:103:	
GTTCTGATTG CTTTTATTTC C	21
(2) INFORMATION FOR SEQ ID NO:104:	
(i) SEQUENCE CHARACTERISTICS:	

6,033,857

159

160

-continued

(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:104:

ATCACTTCTT CCATTGCATC 20

(2) INFORMATION FOR SEQ ID NO:105:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:105:

CCGTGGCTGG TAAATCTG 18

(2) INFORMATION FOR SEQ ID NO:106:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:106:

CTGGTAGCTC CAACTAATC 19

(2) INFORMATION FOR SEQ ID NO:107:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

6,033,857

161

162

-continued

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:107:
ACCGGTACAA ACCTTTCATT G 21

(2) INFORMATION FOR SEQ ID NO:108:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:108:
CTATTTTGAT TTGCTTTTAT TATT 24

(2) INFORMATION FOR SEQ ID NO:109:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:109:
GCTATTTCTT TGATACTGGA C 21

(2) INFORMATION FOR SEQ ID NO:110:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:110:
TTGGAACAT AAATATGTGG G 21

(2) INFORMATION FOR SEQ ID NO:111:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

6,033,857

163

164

-continued

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:111:

ACTTACAGGA GCCACATAAC 20

(2) INFORMATION FOR SEQ ID NO:112:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:112:

CTACATTAAT TATGATAGGC TCG 23

(2) INFORMATION FOR SEQ ID NO:113:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:113:

GTACTAATGT GTGGTTTGAA A 21

(2) INFORMATION FOR SEQ ID NO:114:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:114:

TCAATGCAAG TTCTTCGTCA GC 22

6,033,857

165

166

-continued

(2) INFORMATION FOR SEO ID NO:115:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 21 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

```
(ii) MOLECULE TYPE: other nucleic acid
      (A) DESCRIPTION: /desc = "Primer"
```

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:115:

GGGAAGCTTC ATAAGTCAGT C

21

(2) INFORMATION FOR SEO ID NO:116:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

```
(ii) MOLECULE TYPE: other nucleic acid
      (A) DESCRIPTION: /desc = "Primer"
```

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:116:

TTTGTAAATGA AGCATCTGAT ACC

23

(2) INFORMATION FOR SEO ID NO:117:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

```
(ii) MOLECULE TYPE: other nucleic acid
      (A) DESCRIPTION: /desc = "Primer"
```

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:117:

AATGATGAAT GTAGCACGC

19

(2) INFORMATION FOR SEQ ID NO:118:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

```
(ii) MOLECULE TYPE: other nucleic acid
      (A) DESCRIPTION: /desc = "Primer"
```

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:118:

6,033,857

167

168

-continued

GTCTGAATGT TCGTTACT

18

(2) INFORMATION FOR SEO ID NO:119:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

```
(ii) MOLECULE TYPE: other nucleic acid
      (A) DESCRIPTION: /desc = "Primer"
```

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:119:

ACCATCAAAC ACATCATCC

19

(2) INFORMATION FOR SEO ID NO:120:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

```
(ii) MOLECULE TYPE: other nucleic acid
      (A) DESCRIPTION: /desc = "Primer"
```

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:120:

AGAAAGTAAC TTGGAGGGAG

20

(2) INFORMATION FOR SEQ ID NO:121:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

```
(ii) MOLECULE TYPE: other nucleic acid
      (A) DESCRIPTION: /desc = "Primer"
```

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:121:

CTCCTGAAAC TGTTCCTTG G

21

(2) INFORMATION FOR SEO ID NO:122:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 21 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

```
(ii) MOLECULE TYPE: other nucleic acid
      (A) DESCRIPTION: /desc = "Primer"
```

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:122:

6,033,857

169

170

-continued

TAATGGTGCT GGGATATTG G

21

(2) INFORMATION FOR SEQ ID NO:123:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

```
(ii) MOLECULE TYPE: other nucleic acid
      (A) DESCRIPTION: /desc = "Primer"
```

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEO ID NO:123:

GAATGTCGAA GAGCTTGTC

19

(2) INFORMATION FOR SEQ ID NO:124:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

```
(ii) MOLECULE TYPE: other nucleic acid
      (A) DESCRIPTION: /desc = "Primer"
```

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEO ID NO:124:

AAACATACGC TTAGCCAGAC

20

What is claimed is:

1. A method for identifying a mutant BRCA2 nucleotide sequence in a suspected mutant BRCA2 allele which comprises comparing the nucleotide sequence of the suspected mutant BRCA2 allele with the wild-type BRCA2 nucleotide sequence, wherein a difference between the suspected mutant and the wild-type sequences identifies a mutant BRCA2 nucleotide sequence.

2. A method for diagnosing a predisposition for breast cancer in a human subject which comprises comparing the germline sequence of the BRCA2 gene or the sequence of its mRNA in a tissue sample from said subject with the germline sequence of the wild-type BRCA2 gene or the sequence of its mRNA, wherein an alteration in the germline sequence of the BRCA2 gene or the sequence of its mRNA of the subject indicates a predisposition to said cancer.

3. The method of claim 2 wherein an alteration is detected in a regulatory region of the BRCA2 gene.

4. The method of claim 2 wherein the detection in the alteration in the germline sequence is determined by an assay selected from the group consisting of (a) observing shifts in electrophoretic mobility of single-stranded DNA on non-denaturing polyacrylamide gels, (b) hybridizing a BRCA2 gene probe to genomic DNA isolated from said tissue sample, (c) hybridizing an allele-specific probe to genomic DNA of the tissue sample, (d) amplifying all or part of the BRCA2 gene from said tissue sample to produce an amplified sequence and sequencing the amplified sequence, (e) amplifying all or part of the BRCA2 gene from said

tissue sample using primers for a specific BRCA2 mutant allele, (f) molecularly cloning all or part of the BRCA2 gene from said tissue sample to produce a cloned sequence and sequencing the cloned sequence, (g) identifying a mismatch between (1) a BRCA2 gene or a BRCA2 mRNA isolated from said tissue sample, and (2) a nucleic acid probe complementary to the human wild-type BRCA2 gene sequence, when molecules (1) and (2) are hybridized to each other to form a duplex, (h) amplification of BRCA2 gene sequences in said tissue sample and hybridization of the amplified sequences to nucleic acid probes which comprise wild-type BRCA2 gene sequences, (i) amplification of BRCA2 gene sequences in said tissue sample and hybridization of the amplified sequences to nucleic acid probes which comprise mutant BRCA2 gene sequences, (j) screening for a deletion mutation in said tissue sample, (k) screening for a point mutation in said tissue sample, (l) screening for an insertion mutation in said tissue sample, (m) in situ hybridization of the BRCA2 gene of said tissue sample with nucleic acid probes which comprise the BRCA2 gene.

5. A method for detecting a mutation in a neoplastic lesion at the BRCA2 gene in a human subject which comprises comparing the sequence of the BRCA2 gene or the sequence of its mRNA in a tissue sample from a lesion of said subject with the sequence of the wild-type BRCA2 gene or the sequence of its mRNA, wherein an alteration in the sequence of the BRCA2 gene or the sequence of its mRNA of the subject indicates a mutation at the BRCA2 gene of the neoplastic lesion.

6,033,857

171

6. The method of claim 5 wherein an alteration is detected in the a regulatory regions of the BRCA2 gene.
7. The method of claim 5 wherein the detection in the alteration in the BRCA2 sequence is determined by an assay selected from the group consisting of (a) observing shifts in electrophoretic mobility of single-stranded DNA on non-denaturing polyacrylamide gels, (b) hybridizing a BRCA2 gene probe to DNA isolated from said tissue sample, (c) hybridizing an allele-specific probe to DNA of the tissue sample, (d) amplifying all or part of the BRCA2 gene from said tissue sample to produce an amplified sequence and sequencing the amplified sequence, (e) amplifying all or part of the BRCA2 gene from said tissue sample using primers for a specific BRCA2 mutant allele, (f) molecularly cloning all or part of the BRCA2 gene from said tissue sample to produce a cloned sequence and sequencing the cloned sequence, (g) identifying a mismatch between (1) a BRCA2 gene or a BRCA2 mRNA isolated from said tissue sample, and (2) a nucleic acid probe complementary to the human wild-type BRCA2 gene sequence, when molecules (1) and (2) are hybridized to each other to form a duplex, (h) amplification of BRCA2 gene sequences in said tissue

172

sample and hybridization of the amplified sequences to nucleic acid probes which comprise wild-type BRCA2 gene sequences, (i) amplification of BRCA2 gene sequences in said tissue sample and hybridization of the amplified sequences to nucleic acid probes which comprise mutant BRCA2 gene sequences., (0) screening for a deletion mutation in said tissue sample, (k) screening for a point mutation in said tissue sample, (l) screening for an insertion mutation in said tissue sample, (m) in situ hybridization of the BRCA2 gene of said tissue sample with nucleic acid probes which comprise the BRCA2 gene.
8. A method for confirming the lack of a BRCA2 mutation in a neoplastic lesion from a human subject which comprises comparing the sequence of the BRCA2 gene or the sequence of its mRNA in a tissue sample from a lesion of said subject with the sequence of the wild-type BRCA2 gene or the sequence of its RNA, wherein the presence of the wild-type sequence in the tissue sample indicates the lack of a mutation at the BRCA2 gene.

* * * * *

UNITED STATES PATENT AND TRADEMARK OFFICE
CERTIFICATE OF CORRECTION

PATENT NO. : 6,033,857
APPLICATION NO. : 09/044946
DATED : March 7, 2000
INVENTOR(S) : Tavgian et al.

Page 1 of 1

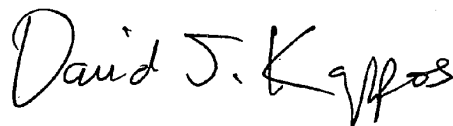
It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

Col. 172, Claim 7, line 6, please change "(0)" to --(j)--.

Col. 172, Claim 8, line 17, please change "RNA" to --mRNA--.

Signed and Sealed this

Twenty-fourth Day of November, 2009



David J. Kappos
Director of the United States Patent and Trademark Office

A000544



US006951721B2

**(12) United States Patent
Murphy****(10) Patent No.: US 6,951,721 B2
(45) Date of Patent: *Oct. 4, 2005****(54) METHOD FOR DETERMINING THE
HAPLOTYPE OF A HUMAN BRCA1 GENE****(75) Inventor: Patricia D. Murphy**, Slingerlands, NY
(US)**(73) Assignee: Gene Logic Inc.**, Gaithersburg, MD
(US)**(*) Notice:** Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 87 days.

This patent is subject to a terminal disclaimer.

(21) Appl. No.: 09/923,327**(22) Filed: Aug. 8, 2001****(65) Prior Publication Data**

US 2003/0096236 A1 May 22, 2003

Related U.S. Application Data**(63)** Continuation-in-part of application No. 09/084,471, filed on May 22, 1998, and a continuation-in-part of application No. 08/905,772, filed on Aug. 4, 1997, now abandoned, which is a continuation-in-part of application No. 08/798,691, filed on Feb. 12, 1997, now Pat. No. 5,750,400, which is a continuation-in-part of application No. 08/598,591, filed on Feb. 12, 1996, now Pat. No. 5,654,155.**(51) Int. Cl.⁷ C12Q 1/68; C12P 19/34;
C07H 21/02; C07H 21/04****(52) U.S. Cl. 435/6; 435/91.1; 435/91.2;
536/23.1; 536/23.5; 536/24.31****(58) Field of Search 435/6, 91.1, 91.2;
536/23.1, 23.5, 24.31; 530/300, 350, 386****(56) References Cited****U.S. PATENT DOCUMENTS**

4,458,066 A	7/1984	Caruthers et al.	536/27
5,498,324 A	3/1996	Yeung et al.	204/452
5,561,058 A	10/1996	Gelfand et al.	435/912
5,578,443 A *	11/1996	Santamaria et al.	435/6
5,593,840 A	1/1997	Bhatnagar et al.	435/6
5,642,936 A	7/1997	Evans	128/630
5,654,155 A *	8/1997	Murphy et al.	435/6
5,686,246 A	11/1997	Kornman et al.	435/6
5,750,400 A *	5/1998	Murphy et al.	435/6
5,763,183 A *	6/1998	Pesonen et al.	435/6

FOREIGN PATENT DOCUMENTS

EP	0 699 754 A1	3/1996
EP	0 705 902 A1	4/1996
EP	0 705 903 A1	4/1996
WO	WO 95/19369	7/1995
WO	WO 98/05677	2/1998
WO	WO 98/44157	10/1998

OTHER PUBLICATIONSBeaucage et al., "Deoxynucleoside Phosphoramidites—A New Class of Key Intermediates for Deoxypolynucleotide Synthesis," *Tetrahedron Letters* 22(20):1859–1862 (1981).Beutler et al., "Molecular Heterogeneity of Glucose-6-Phosphate Dehydrogenase A," *Blood* 74(7):2550–2555 (1989).Beutler et al., "Mutation analysis of glucose-6-phosphate dehydrogenase (G6PD) variants in Costa Rica," *Human Genetics* 87:462–464 (1991).Connor et al., "Detection of sickle cell β^S -globin allele by hybridization with synthetic oligonucleotides," *Proc. Natl. Acad. Sci. U.S.A.* 80:278–282 (1983).Cooke et al., "P^{1A2} polymorphism and efficacy of aspirin," *The Lancet* 351:1253 (1998).Couch et al., "Mutations and Polymorphisms in the Familial Early-Onset Breast Cancer (BRCA1) Gene," *Human Mutation* 8:8–18 (1996).Dunning et al., "Common BRCA1 variants and susceptibility to breast and ovarian cancer in the general population," *Human Molecular Genetics* 6(2):285–289 (1997).Durocher et al., "Comparison of BRCA1 polymorphisms, rare sequence variants and/or missense mutations in unaffected and breast/ovarian cancer populations," *Human Molecular Genetics* 5(6):835–842 (1996).Easton et al., "Genetic Linkage Analysis in Familial Breast and Ovarian Cancer: Results from 214 Families," *American Journal of Human Genetics* 52:678–701 (1993).Friedman et al., "Confirmation of BRCA1 by analysis of germline mutations linked to breast and ovarian cancer in ten families," *Nature Genetics* 8:399–404 (1994).Friend et al., Breast cancer information on the web, *Nature Genetics* 11:238–239 (1995).Hirono, A. and Beutler, E., "Molecular cloning and nucleotide sequence of cDNA for human glucose-6-phosphate dehydrogenase variant A(-)," *Proc. Natl. Acad. Sci. U.S.A.* 85:3951–3954 (1988).Holt et al., "Growth retardation and tumour inhibition by BRCA1," *Nature Genetics* 12(3):298–302 (1996).Landgren et al., "A Ligase-Mediated Gene Detection Technique," *Science* 241:1077–1021 (1988).Landgren et al., "DNA Diagnostics—Molecular Techniques and Automation," *Science* 242:229–237 (1988).Liaw et al., "Germline mutations of the PTEN gene in Cowden disease, an inherited breast and thyroid cancer syndrome," *Nature Genetics* 16(1):64–67 (1997).

(Continued)

Primary Examiner—Carla J. Myers**(74) Attorney, Agent, or Firm**—Morgan, Lewis & Bockius LLP**(57) ABSTRACT**

Methods for identifying functional allele profiles of a given gene are disclosed. Functional allele profiles comprise the commonly occurring alleles in a population, and the relative frequencies at which such alleles of a given gene occur. Functional allele profiles are useful in treatment and diagnosis of diseases, for genetic and pharmacogenetic applications and for evaluating the degree to which the gene(s) are under selective pressure.

19 Claims, 1 Drawing Sheet**A000545**

US 6,951,721 B2

Page 2

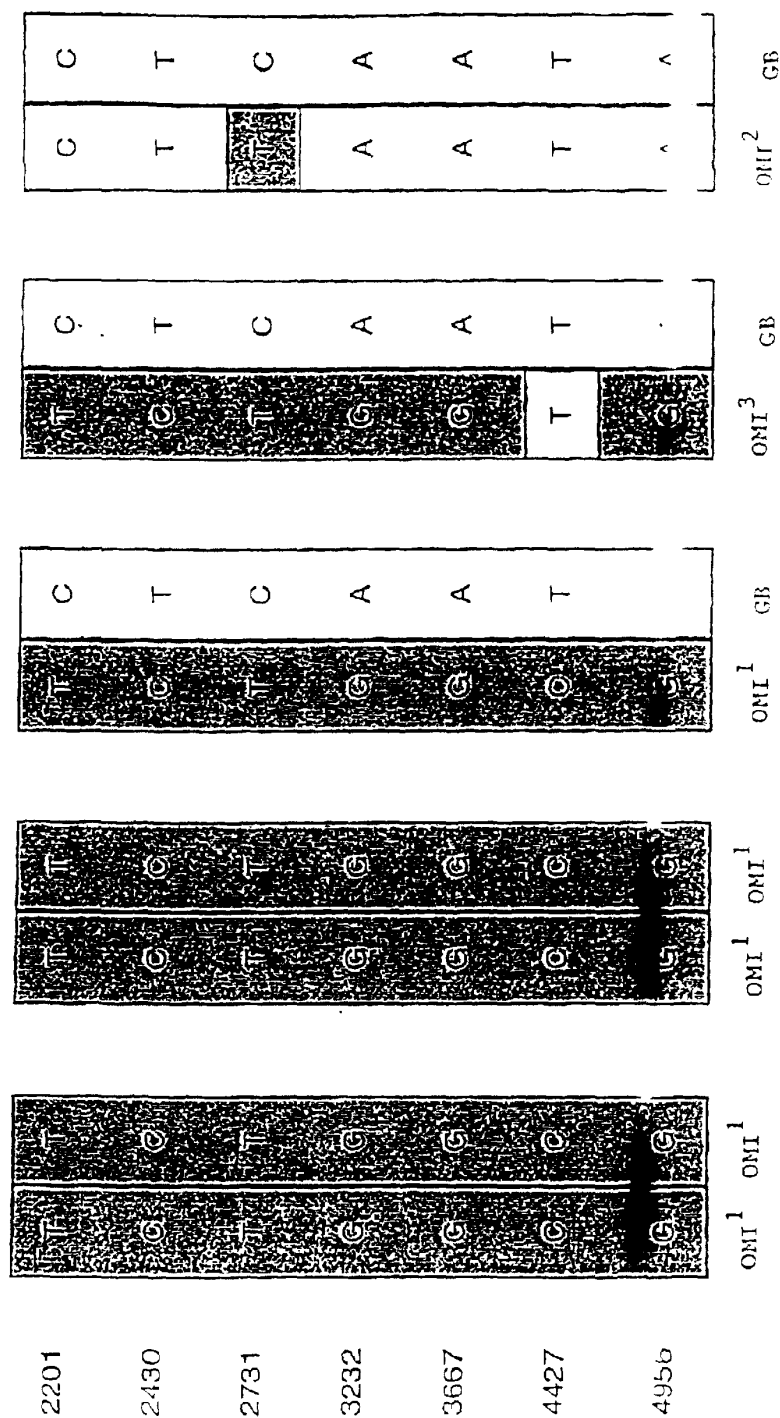
OTHER PUBLICATIONS

- Maniatis et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, NY, pp. 280–281 (1982).
- Miki et al., “A Strong Candidate for the Breast and Ovarian Cancer Susceptibility Gene BRCA1,” *Science* 266:66–71 (1994).
- Newman et al., “The Human Platelet Alloantigens, P1^{A1} and P1^{A2} Are Associated with a Leucine³³/Proline³³ Amino Acid Polymorphism in Membrane Glycoprotein IIIa, and Are Distinguishable by DNA Typing,” *Journal of Clinical Investigation* 83:1778–1781 (1989).
- PCR, *A Practical Approach*, IRL Press, Eds. M.J. McPherson, P. Quirke, and G.R. Taylor (1992).
- Ridker et al., “P1^{A1/A2} polymorphism of platelet glycoprotein IIIa and risks of myocardial infarction, stroke, and venous thrombosis,” *The Lancet* 349:385–388 (1997).
- Saiki et al., “A Novel Method for the Detection of Polymorphic Restriction Sites by Cleavage of Oligonucleotide Probes: Application to Sickle-Cell Anemia,” *Bio/Technology* 3:1008–1012 (1985).
- Sambrook et al., *Molecular Cloning, A Laboratory Manual*, 2nd Ed.; Cold Spring Harbor Laboratory Press, at 9.16–9.19 (1989).
- Sharan et al., “Murine Brce1: sequence and significance for human missense mutations,” *Human Molecular Genetics* 4(12):2275–2278 (1995).
- Shuldiner et al., *Handbook of Endocrine Research Techniques*, de Pablo, F., Scanes, C., and Weintraub, B., eds., Academic Press, Inc., pp. 457–486 (1993).
- Storey et al., “Role of a p53 polymorphism in the development of human papilloma-virus-associated cancer,” *Nature* 393:229–234 (1998).
- Vogelstein et al., “p53 Function and Dysfunction,” *Cell* 70:523–526 (1992).
- Weiss et al., “A monoclonal antibody (SZ21) specific for platelet GPIIIa distinguishes P1^{A1} from P1^{A2},” *Tissue Antigens* 46:374–381 (1995).
- Weiss et al., “A Polymorphism of a Platelet Glycoprotein Receptor as an Inherited Risk Factor for Coronary Thrombosis,” *The New England Journal of Medicine* 334(17):1090–1094 (1996).
- Molecular Biology, 2d ed., David Freifelder, auth., Jones and Bartlett Publishers, Inc. Boston, p. 37 (1987).
- Mattila, P. et al., “Extensive Allelic Sequence Variation in the J Region of the Human Immunoglobulin Heavy Chain Gene Locus,” *Eur. J. Immunol.*—25:2578–2582 (1995).
- Morrison, N., et al., “Frequent Alleles of the Human Vitamin D Receptor Gene Are Functionally Distinct,” *Garvan Institute of Medical Research*—20 (Feb. 21, 1992).
- Moudallal, J., et al., “Monoclonal Antibodies as Probes of the Antigenic Structure of Tobacco Mosaic Virus,” *The EMBO Journal*—1(8):1005–1010 (1982).
- Hacia, J., et al., “Detection of Heterozygous Mutations in BRCA1 Using High Density Oligonucleotide Arrays and Two-Colour Fluorescence Analysis,” *Nature Genetics* 14:441–447 (Dec. 1996).
- Merajver, S., et al., “Risk Assessment and Presymptomatic Molecular Diagnosis in Hereditary Breast Cancer,” *DNA Technology*—16(1):139–137 (Mar. 1996).
- Madsen, H., et al., “A New Frequent Allele is the Missing Link in the Structural Polymorphism of the Human Mannan-Binding Protein,” *Immunogenetics*—40:37–44 (1994).
- Wainscoat, J., et al., “Evolutionary Relationships of Human Populations From an Analysis of Nuclear DNA Polymorphisms,” *Nature*—319:491–493 (Feb. 6, 1986).

* cited by examiner

A000546

FIGURE 1



POSITION

US 6,951,721 B2

1

METHOD FOR DETERMINING THE HAPLOTYPE OF A HUMAN BRCA1 GENE

This application is a continuation-in-part of U.S. application Ser. No. 08/905,772 filed Aug. 4, 1997 now abandoned, which is a continuation-in-part of U.S. application Ser. No. 08/798,691 filed Feb. 12, 1997 now U.S. Pat. No. 5,750,400 issued May 12, 1998, which is a continuation-in-part of U.S. application Ser. No. 08/598,591 filed now Feb. 12, 1996 U.S. Pat. No. 5,654,155 issued Aug. 5, 1997. This application is also a continuation-in-part of U.S. application Ser. No. 09/084,471 filed May 22, 1998, each of which is hereby incorporated by reference in its entirety.

FIELD OF THE INVENTION

The invention relates to methods for identifying functional alleles commonly occurring in a population, for finding new functional alleles, for determining the relative frequencies at which such alleles, for genetic and pharmacogenetic applications of the methods and products produced thereby.

BACKGROUND OF THE INVENTION

An increasing number of genes which play a role in many different diseases are being identified. Detection of mutations in such genes is instrumental in determining susceptibility to or diagnosing these diseases. Some diseases, such as sickle cell disease, are known to be monomorphic; i.e., the disease is generally caused by a single mutation present in the population. In such cases where one or only a few known mutations are responsible for the disease, methods for detecting the mutations are targeted to the site within the gene at which they are known to occur. However, the mutation responsible for such a monomorphic disease can only be established in the first instance if there exists an accurate reference sequence for the non-pathological state.

In many other cases individuals affected by a given disease display extensive allelic heterogeneity. For example, more than 125 mutations in the human BRCA1 gene have been reported (Breast Cancer Information Core world wide web site at http://www.nchgr.nih.gov/dir/lab_transfer/bic, which became publicly available on Nov. 1, 1995; Friend, S. et al., 1995, Nature Genetics 11:238). Mutations in the BRCA1 gene are thought to account for roughly 45% of inherited breast cancer and 80–90% of families with increased risk of early onset breast and ovarian cancer (Easton, 1993, et al., American Journal of Human Genetics 52: 678–701).

Other examples of genes for which the population displays extensive allelic heterogeneity and which have been implicated in disease include CFTR (cystic fibrosis), dystrophin (Duchenne muscular dystrophy, and Becker muscular dystrophy), and p53 (Li-Fraumeni syndrome).

Breast cancer is also an example of a disease in which, in addition to allelic heterogeneity, there is genetic heterogeneity. In addition to BRCA1, the BRCA2 and BRCA3 genes have been linked to breast cancer. Similarly, the NF1 and NF2 genes are involved in neurofibromatosis (types I and II, respectively). Furthermore, hereditary non-polyposis colorectal cancer (HNPCC) is a disease in which four genes, MSH2, MLH1, PMS1, and PMS2, have been implicated. It is yet another example of a disease in which there is both allelic and genetic heterogeneity of mutations. A cDNA sequence for MSH2 has been deposited in GenBank as Accession No. U03911; and a cDNA sequence for MLH1 has been deposited in GenBank as Accession No. U40978.

2

Additionally, disease or disease susceptibility also results from the interaction of more than one gene or the interaction of an environmental, chemical or biological influence on one or more genes. For example, measles virus infects many people; some are immune due to vaccination or previous infection, some are infected but asymptomatic, some become sick with a rash, some develop an encephalitis and some die. Genetic susceptibility and many other factors are involved in the outcome.

A common misconception in the field of molecular genetics is that for any given gene there exists a single “normal” or “wild-type” sequence. Often, research into such wild-type sequences ends once a single sequence associated with normal function is identified. For example, information in GenBank concerning the BRCA1 sequence represented by GenBank Accession No. U14680 does not indicate a basis for whether this sequence is representative of the population at large. Even when polymorphisms of the BRCA1 gene were identified, no analysis was provided of the arrangement of such sequence variations in a given allele (i.e., the haplotype) (Miki et al., 1994, Science 266: 66–71).

In the fields of plant and animal breeding, the “wild-type” may not be the desirable or may be one of several possibilities. For some domesticated plants and animals, the “wild-type” of any gene may not even be known. In the Brassica family, debate exists as to exactly what is a wild cabbage plant, much less which of the many genes or traits constitutes a “wild-type”. By definition, a wild-type is not pathological but sometimes this definition seems inappropriate. For example, the MacIntosh apple is propagated asexually exclusively. An inability to reproduce naturally may be considered the result of pathological mutation(s) but is none the less the desired trait. In other situations, different strains of a plant are cross-bred where each set of genes from each parent strain may be considered “wild-type”.

Identification of a mutation provides for early diagnosis which is essential for effective treatment of many diseases. However, in order to identify a mutation, it is necessary to have an accurate understanding of the proper reference sequences which encode the non-pathological functional gene products occurring in the population. Prior research efforts and publications have neither suggested nor taught a systematic approach to both identify a functional allele of a given gene and determine the relative frequency with which the allele occurs in the population.

Certain wild-type sequences of a gene may be otherwise indistinguishable from others except under certain circumstances. For example, a gene involved in resistance or susceptibility to a certain infectious agent is only recognized when the individual plant or animal is exposed to the infectious agent. Likewise chemical sensitivity may be a wild-type which is pathological under only certain circumstances which may never occur in the individual. Drought tolerance traits are significant only under environmental stress which may or may not occur. Therefore, the type of wild-type sequence is of importance.

SUMMARY OF THE INVENTION

It is an object of the invention to provide an integrated, systematic process for determining the functional allele profile for a given gene in a population. In accordance with the invention, a functional allele profile contains 1) the identity of the key functional allele or alleles for a given gene in the population, including the “consensus” sequence, and 2) the relative frequency with which these functional alleles occur in the population. Thus, the functional allele

A000548

US 6,951,721 B2

3

profile includes the identification of the consensus normal sequence, i.e., the most commonly occurring functional allele.

The present invention, therefore, provides a normal sequence which is the most likely sequence to be found in the majority of the normal population, the (i.e., "consensus normal DNA sequence"). A consensus normal allele sequence of a gene more accurately reflects the most likely sequence to be found in the population. Determining the consensus sequence is useful in both the diagnosis and treatment of disease. For example, use of the consensus normal gene sequence reduces the likelihood of misinterpreting a "sequence variation" found in the normal population with a pathologic "mutation" (i.e. causes disease in the individual or puts the individual at a high risk of developing the disease). A consensus normal DNA sequence makes it possible for true pathological mutations to be easily identified or differentiated from polymorphisms.

With large interest in mutation and polymorphism testing such as cancer predisposition testing, misinterpretation of sequence data is a particular concern. Individuals diagnosed with cancer want to know their prognosis and whether their disease is caused by a heritable genetic mutation. Likewise for other disease and traits and those who manage or manipulate these traits. Relatives of those with cancer who have not yet been diagnosed with the disease are also concerned whether they carry such a heritable mutation. Carrying such a mutation may increase risk of contracting the disease sufficiently to warrant an aggressive surveillance program. Accurate and efficient identification of mutations in genes linked to disease is crucial for widespread diagnostic screening for hereditary diseases.

In addition, the consensus sequence, or other sequences identified in the functional allele profile, allow for the selection of therapeutically optimal nucleotide sequences to be administered in gene therapy or gene replacement, or optimal amino acid sequence in the therapeutic administration of active proteins or peptides. The consensus sequence is generally the easiest target for various agonists, antagonists and measuring interactions with the gene or expression product appropriate for pharmacogenetic analysis.

Moreover, determining a functional allele profile of genes allows for an evaluation of the degree to which the gene is under selective pressure.

It is another embodiment of the present invention to find a new allele having a different wild-type haplotype from that previously known.

It is another embodiment of the present invention to determine the haplotype of a sample by determining the polymorphisms constituting the haplotype. Such a technique applies to one and plural genes, especially genes which interact or express products which interact with each other directly, interact with the same or similar other compound or are along the same metabolic pathway. As such, the method of the present invention determines combinations of haplotypes in different genes.

It is another embodiment of the present invention is determining how an individual will react to a particular chemical, environmental or biological influence. It is a premise of the present invention that different wild-type genes or their expression products interact differently in some circumstances.

Another embodiment of the present invention is the determination of traits and susceptibilities of plants and animals during breeding experiments by detecting the polymorphisms constituting the gene haplotype associated with the trait or susceptibility of interest.

4

BRIEF DESCRIPTION OF THE FIGURE

FIG. 1: FIG. 1 shows alternative alleles containing polymorphic (non-mutation causing variations) sites along the BRCA1 gene, represented as individual "haplotypes" of the BRCA1 gene. The alternative allelic variations occurring at nucleotide positions 2201, 2430, 2731, 3232, 3667, 4427, and 4956 are shown. The BRCA1^(omi1) haplotype (SEQ ID NO: 263) is indicated with dark shading. For comparison, the haplotype available in GenBank is completely unshaded and designated as "GB". Two additional haplotypes (BRCA1^(omi2) haplotype (SEQ ID NO: 265) and BRCA1^(omi3) haplotype (SEQ ID NO: 267) are represented with mixed shaded and unshaded positions, numbers 7 and 9 from left to right).

DETAILED DESCRIPTION OF THE INVENTION

The invention provides an integrated, systematic process for determining the functional allele profile for a given gene or combination of genes in a population. In accordance with the invention, a functional allele profile contains 1) the identity of the key functional allele or alleles for a given gene in the population, including the "consensus" sequence, and 2) the relative frequency with which these functional alleles occur in the population. Thus, the functional allele profile includes the identification of the consensus normal sequence, i.e., the most commonly occurring functional allele.

The present invention, therefore, provides a normal sequence which is the most likely sequence to be found in the majority of the normal population, the (i.e., "consensus normal DNA sequence"). A consensus normal allele sequence of a gene more accurately reflects the most likely sequence to be found in the population. In the process for determining functional alleles or afterward, one may search for and discover or synthesize a heretofor unknown or "new" allele.

A functional allele profile can be determined for any gene in which an altered or deficient function produces a recognizable, phenotypic trait, including, but not limited to, pathology. The invention is set forth for the purpose of illustration, and not by way of limitation, for determining the functional allele profile of three different genes associated with disease—for example, the MSH2 and MLH1 genes, each associated with hereditary non-polyposis colorectal cancer (HNPCC), and the BRCA1 gene, associated with breast, ovarian, prostate and other cancers.

The following terms as used herein are defined as follows:

"Allele" refers to an alternative version (i.e., nucleotide sequence) of a gene or DNA sequence at a specific chromosomal locus.

"Allelic variation" or "sequence variation" refers to a particular alternative nucleotide or nucleotide sequence at a position within a gene (e.g., a polymorphic site or mutation) whose sequence varies from one allele to another.

"Coding sequence" or "DNA coding sequence" refers to those portions of a gene which, taken together, code for a peptide (protein), or which nucleic acid itself has function.

"Composite genomic sequence" refers to the combination of the two allelic nucleotide sequences (i.e., maternal and paternal) obtained from sequencing a diploid genomic sample.

"Consensus" refers to the most commonly occurring in the population.

"Functional allele" refers to an allele which is naturally transcribed and translated into a functioning protein.

A000549

US 6,951,721 B2

5

“Functional Allele Profile” refers to a set of functional alleles which are representative of the most common alleles occurring in a population, wherein the functional alleles are identified by nucleotide sequence and the relative frequencies with which the functional alleles occur in the population.

“Haplotype” refers to a set of nucleotides or nucleotide sequences occurring at sites of allelic variation occurring within a locus on a single chromosome (of either maternal or paternal origin). The “locus” includes the entire coding sequence.

“Mutation” refers to a base change or a gain or loss of base pair(s) in a DNA sequence, which results in a DNA sequence which codes for a non-functioning protein or a protein with substantially reduced or altered function.

“Agent for polymerization” refers to an enzyme which may be heat stable, e.g. Taq polymerase, or function at lower temperatures, e.g., room temperature, that effects an extension of DNA from a short primer sequence annealed to the target DNA of interest.

“Polymorphism” refers to an allelic variation which occurs in greater than or equal to 1% of the normal healthy population.

“Single nucleotide polymorphism” (SNP) refers to an allelic variation which is defined by two (and only two) alternative bases found at a specific and particular nucleotide in genomic DNA. It may be within a gene (i.e., exonic or intronic) or outside of a gene (such as in a promoter or other regulatory structure) or lastly found between genes.

“Individual” refers to a single organism which may be human, plant or non-human animal. The individual may be intact or a biological sample taken from the individual which contains sufficient substances or information regarding the individual.

“Protein variant” and “variant amino acid sequence” refers to different amino acid sequences from that in one naturally occurring wild-type protein and is generally considered the same protein. Some different haplotypes have variant amino acid sequences.

“Expression product” refers to an RNA, spliced or unspliced, a pre-, pro-, prepro- or a peptide which alone or in conjunction with other peptides constitutes a protein.

“Pharmaceutical” refers to any bio-affecting chemical drug or biological agent which alters or induces an alteration in the metabolism of an “individual”. Pharmaceuticals include compositions for use on veterinary animals and agricultural and ornamental plants.

“Trait” refers to a phenotypically determinable characteristic resulting from the influence of one or more genes, alone or in conjunction with an environmental condition or exposure to other agents. Traits include susceptibilities to chemicals, infectious agents and environmental conditions (temperature, drought etc.).

Utility of the Invention

A person skilled in the art of genetic testing will find the present invention useful for diagnosis and treatment of diseases and susceptibility thereto. The invention is especially useful for establishing the “standard” (i.e., consensus normal DNA sequence) and new haplotypes for clinical diagnostic, therapeutic, genetic testing and breeding uses.

Diagnostics

The diagnostic applications for which determining a functional allele profile in accordance with the invention include, but are not limited to, the following:

- a) identifying individuals having a gene with no coding mutations, which individuals are therefore not at risk or

6

have no increased susceptibility to the pathology(s) associated with a mutation in the gene in question;

- b) avoiding misinterpretation of functional polymorphisms detected in the gene as mutations;
- c) identifying individuals having a potentially abnormal gene that does not match the Consensus Normal DNA sequence;
- d) determining ethnic founder haplotypes so that clinical analysis is appropriate for an individual from this ethnic group;
- e) determining a sequence under strongest selective pressure; and
- f) determining an amino acid and/or short nucleic acid sequence which may be derived from the consensus normal DNA sequence to make diagnostic and probes antibodies. Labeled diagnostic probes may be used by any hybridization method to determine the level of protein in serum or lysed cell suspension of a patient, or solid surface cell sample such as for immunohistochemical analysis.
- g) detecting a new haplotype and determining the polymorphisms constituting the new haplotype.
- h) detecting a new protein variant type and determining the variant amino acids constituting the new protein variant.
- i) determining the combination of one haplotype or polymorphism for one gene and the haplotype or polymorphism for another different gene in the same individual. Generally, the genes or their expression products interact with each other directly, e.g. bind to each other, or indirectly by functioning with each other on the same substrate, are in different stages in a metabolic pathway, or are related to the same disease, susceptibility, condition or trait.
- j) determining whether to administer a bioeffecting composition to an individual wherein individuals with different haplotypes for one or more genes respond differently to the composition.
- k) determining susceptibility to disease or other pathology to decide on prophylaxis, therapy or differential monitoring.
- l) determining a trait by quick assay of a genetic engineered or selectively bred individual. This permits one to determine the trait without actually measuring the trait phenotypically.
- m) developing probe chips and panels of allele-specific oligonucleotide(s) to assay for the haplotypes or polymorphisms in one or more genes.

Therapeutics

Certain “normal” alleles may be more functional or hyper-functional than the minimum needed to maintain a normal phenotype in an individual, particularly when stressed. By determining the most common allele in a population one may be observing empiric data for such suitability for survival (the effects may be so subtle that scientists have not determined the basis of this selection). For example, alleles with longer mRNA or protein half-lives (i.e., stability) may produce healthier cells, and, thus, healthier people. Conversely, there may also be a selective advantage to a very short RNA half-life such as in proteins involved in the cell cycle pathway. Furthermore, proteases are known to have favored cutting sites which may be present or absent in different normal alleles leading to peptides that have intrinsic activity themselves.

A000550

US 6,951,721 B2

7

Thus the determination of the functional allele profile or a new functional allele in accordance with the invention is useful in clinical therapy for:

- a) selecting optimal alleles for performing gene repair or gene therapy; and
- b) selecting optimal amino acid sequence for administration of functional protein in treatment or prevention of diseases.

Evolution and Population Genetics Analysis

The determination of the functional allele profile or a new functional allele in accordance with the invention is useful for:

- a) determining whether a particular gene is under strong selective pressure; and
- b) determining which of two or more genes which encode proteins with similar functions represents a redundant, or back-up copy of the gene.

Stepwise Process for Determining Functional Allele Profile

For the purpose of illustration, and not by way of limitation, the invention is described below for determining the functional allele profile of three cancer genes. However,

8

In accordance with the invention, a group of individuals determined to be at low risk for carrying a mutation in the gene of interest is used as a source for genetic material. Any standard method known in the art for performing pedigree analysis can be used for this selection process. See, for example, Harper, P. S., *Practical Genetic Counseling*, 3d. ed., 1988 (Wright/Butterworth & Co. Ltd.: Boston), especially at pages 4–7. For example, individuals can be screened in order to identify those with no disease history in their immediate family, i.e., among their first and second degree relatives. A first degree relative is a parent, sibling, or offspring. A second degree relative is an aunt, uncle, grandparent, grandchild, niece, nephew, or half-sibling.

In a preferred embodiment for when a functional allele profile of an autosomal dominant disorder with relatively high penetrance (e.g., greater than 50%) is desired, each person is asked to fill out a hereditary cancer prescreening questionnaire. More preferably, when an autosomal dominant cancer gene with such relatively high penetrance is the gene of interest, the questionnaire set forth in Table 1, below, is used.

TABLE 1

Hereditary Cancer Pre-Screening Questionnaire	
Part A: Answer the following questions about your family	
1. To your knowledge, has anyone in your family been diagnosed with a very specific hereditary colon disease called Familial Adenomatous Polyposis (FAP)?	
2. To your knowledge, have you or any aunt had breast cancer diagnosed before the age 35?	
3. Have you had Inflammatory Bowel Disease, also called Crohn's Disease or Ulcerative Colitis, for <u>more</u> than 7 years?	
Part B: Refer to the list of cancers below for your responses only to questions in Part B	
Bladder Cancer, Lung Cancer, Pancreatic Cancer, Breast Cancer, Gastric Cancer, Prostate Cancer, Colon Cancer, Malignant Melanoma, Renal Cancer, Endometrial Cancer, Ovarian Cancer, Thyroid Cancer	
4. Have your mother or father, your sisters or brothers or your children had any of the listed cancers?	
5. Have there been diagnosed in your <u>mother's</u> brothers or sisters, or your <u>mother's</u> parents <u>more than one</u> of the cancers in the above list?	
6. Have there been diagnosed in your <u>father's</u> brothers or sisters, or your <u>father's</u> parents <u>more than one</u> of the cancers in the above list?	
Part C: Refer to the list of relatives below for responses only to questions in Part C	
You, Your mother, Your sisters or brothers, Your mothers's sisters or brothers (maternal aunts and uncles), Your children, Your mother's parents (maternal grandparents)	
7. Have there been diagnosed in these relatives <u>2 or more identical</u> types of cancer? Do not count "simple" skin cancer, also called basal cell or squamous cell skin cancer.	
8. Is there a <u>total of 4 or more</u> of any cancers in the list of relatives above other than "simple" skin cancers?	
Part D: Refer to the list of relatives below for responses only to questions in Part D.	
You, Your father, Your sisters or brothers, Your fathers's sisters or brothers (paternal aunts and uncles)	
Your children, Your father's parents (paternal grandparents)	
9. Have there been diagnosed in these relatives <u>2 or more identical</u> types of cancer? Do not count "simple" skin cancer, also called basal cell or squamous cell skin cancer.	
10. Is there a <u>total of 4 or more</u> of any cancers in the list of relatives above other than "simple" skin cancers?	
© Copyright 1996, OncorMed, Inc.	

the same principles can be applied in accordance with the invention to any gene in which a sequence variation results in a phenotypic trait, in any population within any species. 65
Screening for Individuals with Functional Allele Phenotype

Individuals who answer no to all questions in Table 1 are designated as low risk of being carriers of mutations in the gene of interest and, therefore, in accordance with the invention, are candidates for further analysis set forth below.

US 6,951,721 B2

9

Sequencing

From the group of individuals determined to have a low risk of being carriers for a mutant allele of the gene of interest, a group is selected for genomic DNA sequence analysis. Any number of samples may be analyzed. Preferably, a number of samples which is small enough for convenient, accurate sequence analysis, but large enough to provide a reliable representation of the population is analyzed. Most preferably, initial sequencing may be performed on ten different chromosomes by analyzing samples from five unrelated individuals.

Preferably, sequencing template is obtained by amplifying the coding region and optionally one or more related sequences (e.g. splice site junctions, enhancers, introns, promoters and other regulatory elements) of the gene of interest. Any nucleic acid specimen, in purified or non-purified form, can be utilized as the starting nucleic acid or acids, providing it contains, or is suspected of containing, the specific nucleic acid sequence containing a polymorphic locus. Thus, the process may amplify, for example, DNA or RNA, including messenger RNA, wherein DNA or RNA may be single stranded or double stranded. In the event that RNA is to be used as a template, enzymes, and/or conditions optimal for reverse transcribing the template to DNA would be utilized. In addition, a DNA-RNA hybrid which contains one strand of each may be utilized. A mixture of nucleic acids may also be employed, or the nucleic acids produced in a previous amplification reaction herein, using the same or different primers may be so utilized. The specific nucleic acid sequence to be amplified, i.e., the polymorphic locus, may be a fraction of a larger molecule or can be present initially as a discrete molecule, so that the specific sequence constitutes the entire nucleic acid. It is not necessary that the sequence to be amplified be present initially in a pure form; it may be a minor fraction of a complex mixture, such as contained in whole human DNA.

While the primer pairs used are greater than required to amplify the particular polymorphisms, the primer set actually used is listed below. For larger scale testing of polymorphisms for haplotype determination, only the primer pairs actually amplifying the polymorphism are required. Additionally, primers which amplify a shorter region, as short as the one nucleotide polymorphism may be used.

When a gene containing exons is analyzed, preferably the exonic sequences are individually amplified from genomic template DNA using a pair of primers specific for the intronic regions proximally bordering each individual exon.

DNA utilized herein may be extracted from a body sample, such as blood, tissue material and the like by a variety of techniques such as that described by Maniatis, et. al. in *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor, N.Y., pp. 280-281, 1982). If the extracted sample is impure, it may be treated before amplification with an amount of a reagent effective to open the cells, or animal cell membranes of the sample, and to expose and/or separate the strand(s) of the nucleic acid(s). This lysing and nucleic acid denaturing step to expose and separate the strands will allow amplification to occur much more readily.

The deoxyribonucleotide triphosphates dATP, dCTP, dGTP, and dTTP are added to the synthesis mixture, either separately or together with the primers, in adequate amounts and the resulting solution is heated to about 90°-100° C. from about 1 to 10 minutes, preferably from 1 to 4 minutes. After this heating period, the solution is allowed to cool, which is preferable for the primer hybridization. To the cooled mixture is added an appropriate agent for effecting the primer extension reaction (called herein "agent for

10

polymerization"), and the reaction is allowed to occur under conditions known in the art. The agent for polymerization may also be added together with the other reagents if it is heat stable. This synthesis (or amplification) reaction may occur at room temperature up to a temperature above which the agent for polymerization no longer functions. Thus, for example, if DNA polymerase is used as the agent, the temperature is generally no greater than about 40° C. Most conveniently the reaction occurs at room temperature.

The primers used to carry out this invention embrace oligonucleotides of sufficient length and appropriate sequence to provide initiation of polymerization. Environmental conditions conducive to synthesis include the presence of nucleoside triphosphates and an agent for polymerization, such as DNA polymerase, and a suitable temperature and pH. The primer is preferably single stranded for maximum efficiency in amplification, but may be double stranded. If double stranded, the primer is first treated to separate its strands before being used to prepare extension products. The primer must be sufficiently long to prime the synthesis of extension products in the presence of the inducing agent for polymerization. The exact length of primer will depend on many factors, including temperature, buffer, and nucleotide composition. The oligonucleotide primer typically contains 12-20 or more nucleotides, although it may contain fewer nucleotides.

Primers used to carry out this invention are designed to be substantially complementary to each strand of the genomic locus to be amplified. This means that the primers must be sufficiently complementary to hybridize with their respective strands under conditions which allow the agent for polymerization to perform. In other words, the primers should have sufficient complementarity with the 5' and 3' sequences flanking the mutation to hybridize therewith and permit amplification of the genomic locus.

Oligonucleotide primers of the invention are employed in the amplification process which is an enzymatic chain reaction that produces exponential quantities of polymorphic locus relative to the number of reaction steps involved. Typically, one primer is complementary to the negative (-) strand of the polymorphic locus and the other is complementary to the positive (+) strand. Annealing the primers to denatured nucleic acid followed by extension with an enzyme, such as the large fragment of DNA polymerase I (Klenow) and nucleotides, results in newly synthesized + and - strands containing the target polymorphic locus sequence. Because these newly synthesized sequences are also templates, repeated cycles of denaturing, primer annealing, and extension results in exponential production of the region (i.e., the target polymorphic locus sequence) defined by the primers. The product of the chain reaction is a discreet nucleic acid duplex with termini corresponding to the ends of the specific primers employed.

The oligonucleotide primers of the invention may be prepared using any suitable method, such as conventional phosphotriester and phosphodiester methods or automated embodiments thereof. In one such automated embodiment, diethylphosphoramidites are used as starting materials and may be synthesized as described by Beaucage, et al., *Tetrahedron Letters*, 22:1859-1862, (1981). One method for synthesizing oligonucleotides on a modified solid support is described in U.S. Pat. No. 4,458,066.

The agent for polymerization may be any compound or system which will function to accomplish the synthesis of primer extension products, including enzymes. Suitable enzymes for this purpose include, for example, *E. coli* DNA polymerase I, Klenow fragment of *E. coli* DNA polymerase, polymerase muteins, reverse transcriptase, other enzymes,

A000552

US 6,951,721 B2

11

including heat-stable enzymes (i.e., those enzymes which perform primer extension after being subjected to temperatures sufficiently elevated to cause denaturation), such as Taq polymerase. Suitable enzyme will facilitate combination of the nucleotides in the proper manner to form the primer extension products which are complementary to each polymorphic locus nucleic acid strand. Generally, the synthesis will be initiated at the 3' end of each primer and proceed in the 5' direction along the template strand, until synthesis terminates, producing molecules of different lengths.

The newly synthesized strand and its complementary nucleic acid strand will form a double-stranded molecule under hybridizing conditions described above and this hybrid is used in subsequent steps of the process. In the next step, the newly synthesized double-stranded molecule is subjected to denaturing conditions using any of the procedures described above to provide single-stranded molecules.

The steps of denaturing, annealing, and extension product synthesis can be repeated as often as needed to amplify the target polymorphic locus nucleic acid sequence to the extent necessary for detection. The amount of the specific nucleic acid sequence produced will accumulate in an exponential fashion. Amplification is described in *PCR: A Practical Approach*, ILR Press, Eds. M. J. McPherson, P. Quirke, and G. R. Taylor, 1992.

The amplification products may be detected by Southern blots analysis, without using radioactive probes. In such a process, for example, a small sample of DNA containing a very low level of the nucleic acid sequence of the polymorphic locus is amplified, and analyzed via a Southern blotting technique or similarly, using dot blot analysis. The use of non-radioactive probes or labels is facilitated by the high level of the amplified signal. Alternatively, probes used to detect the amplified products can be directly or indirectly detectably labeled, for example, with a radioisotope, a fluorescent compound, a bioluminescent compound, a chemiluminescent compound, a metal chelator or an enzyme. Those of ordinary skill in the art will know of other suitable labels for binding to the probe, or will be able to ascertain such, using routine experimentation.

Sequences amplified by the methods of the invention can be further evaluated, detected, cloned, sequenced, and the like, either in solution or after binding to a solid support, by any method usually applied to the detection of a specific DNA sequence such as PCR, oligomer restriction (Saiki, et al., *BioTechnology*, 3:1008-1012, (1985)), allele-specific oligonucleotide (ASO) probe analysis (Conner, et al., *Proc. Natl. Acad. Sci. U.S.A.*, 80:278, (1983)), oligonucleotide ligation assays (OLAs) (Landgren, et al., *Science*, 241:1007, (1988)), heteroduplex analysis, chromatographic separation and the like. Molecular techniques for DNA analysis have been reviewed (Landgren, et al., *Science*, 242:229-237, (1988)).

Preferably, the method of amplifying is by PCR, as described herein and as is commonly used by those of ordinary skill in the art. Alternative methods of amplification have been described and can also be employed as long as the genetic locus amplified by PCR using primers of the invention is similarly amplified by the alternative means. Such alternative amplification systems include but are not limited to self-sustained sequence replication, which begins with a short sequence of RNA of interest and a T7 promoter. Reverse transcriptase copies the RNA into cDNA and degrades the RNA, followed by reverse transcriptase polymerizing a second strand of DNA. Another nucleic acid amplification technique is nucleic acid sequence-based amplification (NASBA) which uses reverse transcription

12

and T7 RNA polymerase and incorporates two primers to target its cycling scheme. NASBA can begin with either DNA or RNA and finish with either, and amplifies to 10^8 copies within 60 to 90 minutes. Alternatively, nucleic acid can be amplified by ligation activated transcription (LAT). LAT works from a single-stranded template with a single primer that is partially single-stranded and partially double-stranded. Amplification is initiated by ligating a cDNA to the promoter oligonucleotide and within a few hours, amplification is 10^8 to 10^9 fold. Another amplification system useful in the method of the invention is the QB Replicase System. The QB replicase system can be utilized by attaching an RNA sequence called MDV-1 to RNA complementary to a DNA sequence of interest. Upon mixing with a sample, the hybrid RNA finds its complement among the specimen's mRNAs and binds, activating the replicase to copy the tag-along sequence of interest. Another nucleic acid amplification technique, ligase chain reaction (LCR), works by using two differently labeled halves of a sequence of interest which are covalently bonded by ligase in the presence of the contiguous sequence in a sample, forming a new target. The repair chain reaction (RCR) nucleic acid amplification technique uses two complementary and target-specific oligonucleotide probe pairs, thermostable polymerase and ligase, and DNA nucleotides to geometrically amplify targeted sequences. A 2-base gap separates the oligonucleotide probe pairs, and the RCR fills and joins the gap, mimicking DNA repair. Nucleic acid amplification by strand displacement activation (SDA) utilizes a short primer containing a recognition site for HincII with short overhang on the 5' end which binds to target DNA. A DNA polymerase fills in the part of the primer opposite the overhang with sulfur-containing adenine analogs. HincII is added but only cuts the unmodified DNA strand. A DNA polymerase that lacks 5' exonuclease activity enters at the site of the nick and begins to polymerize, displacing the initial primer strand downstream and building a new one which serves as more primer. SDA produces greater than 10^7 -fold amplification in 2 hours at 37° C. Unlike PCR and LCR, SDA does not require instrumented temperature cycling.

Another method is a process for amplifying nucleic acid sequences from a DNA or RNA template which may be purified or may exist in a mixture of nucleic acids. The resulting nucleic acid sequences may be exact copies of the template, or may be modified. The process has advantages over PCR in that it increases the fidelity of copying a specific nucleic acid sequence, and it allows one to more efficiently detect a particular point mutation in a single assay. A target nucleic acid is amplified enzymatically while avoiding strand displacement. Three primers are used. A first primer is complementary to the first end of the target. A second primer is complementary to the second end of the target. A third primer which is similar to the first end of the target and which is substantially complementary to at least a portion of the first primer such that when the third primer is hybridized to the first primer, the position of the third primer complementary to the base at the 5' end of the first primer contains a modification which substantially avoids strand displacement. This method is detailed in U.S. Pat. No. 5,593,840 to Bhatnagar et al., 1997. Although PCR is the preferred method of amplification if the invention, these other methods can also be used to amplify the gene of interest.

A number of methods well-known in the art can be used to carry out the sequencing reactions. Preferably, enzymatic sequencing based on the Sanger dideoxy method is used. Mass spectroscopy may also be used.

The sequencing reactions can be analyzed using methods well-known in the art, such as polyacrylamide gel electro-

A000553

US 6,951,721 B2

13

phoresis. In a preferred embodiment for efficiently processing multiple samples, the sequencing reactions are carried out and analyzed using a fluorescent automated sequencing system such as the Applied Biosystems, Inc. ("ABI", Foster City, Calif.) system. For example, PCR products serving as templates are fluorescently labeled using the Taq Dye Terminator® Kit (Perkin-Elmer cat#401628). Dideoxy DNA sequencing is performed in both forward and reverse directions on an ABI automated Model 377® sequencer. The resulting data can be analyzed using "Sequence Navigator®" software available through ABI.

Alternatively, large numbers of samples can be prepared for and analyzed by capillary electrophoresis, as described, for example, in Yeung et al., U.S. Pat. No. 5,498,324.

Initial and Companion Haplotype Determination

The functional allele profiles identified in accordance with the invention may contain different alleles. Furthermore, each allele may contain multiple allelic variations, such as multiple polymorphisms. In other words, two different alleles may differ in sequence from one another at multiple nucleotide positions. Moreover, two such multiply polymorphic alleles may be present in the same individual, i.e., a heterozygote. When the genomic sample of the gene of such a heterozygous individual is sequenced, the variations at each position can be detected. They are the alternative sequences present at particular positions in the composite sequence obtained from the diploid genome. However, at this stage, which variations are grouped together in each individual haplotype or allele, i.e., the phase of the variations, cannot be determined.

For example, genomic sequence analysis of a hypothetical gene from a heterozygous individual may reveal that polymorphic positions 1, 2, or 3 each contain either an A or a G. However, it cannot be determined from this information alone whether the variations are distributed between the two alleles as:

allele 1=A₁A₂A₃ and allele 2=G₁G₂G₃; or
allele 1=A₁A₂G₃ and allele 2=G₁G₂A₃; or
allele 1=A₁G₂G₃ and allele 2=G₁A₂A₃, etc.

In accordance with the invention, such heterozygous genomic sequences obtained for the purpose of determining a functional allele profile are compared to an initial haplotype sequence. Some haplotypes can also be determined upon sequencing chromosomal samples from a homozygous individual according to the methods above. Such homozygous sequence analyses contain no ambiguities in sequence between the two alleles because they are identical.

Preferably, an initial haplotype sequence is obtained by determining the cDNA sequence of an individual identified as being at low risk for carrying a mutation as described above. Because the full-length of a cDNA of the gene of interest is derived from a single mRNA transcript, it contains the allelic variations of a single haplotype. It contains all of the allelic variations present in a single allele of the individual from which it was obtained. Thus, the cDNA sequence contains half of the allelic variations present in the composite genomic sequence of a heterozygous individual containing that allele. Moreover, unlike sequence information from a heterozygous chromosomal sample, such cDNA sequence indicates which of the allelic variations are grouped together in one allele, i.e., the phase of the variations.

By determining an initial haplotype, the companion haplotype present in a heterozygote can be determined by subtracting this sequence from the composite genomic sequence. For example, if in the illustration set forth above, the cDNA sequenced has an A in position 1, a G in position

14

2 and an A in position 3, then the initial haplotype is A₁G₂A₃. This sequence is then subtracted from the composite genomic sequence to yield the companion haplotype, namely G₁A₂G₃.

In general, the initial haplotype identified in a given individual also can be used to determine the presence of the haplotype in other individuals by comparing the initial haplotype sequence to the composite genomic sequence from such other individuals. When the number of allelic variations detected within a gene is four or greater, and especially when the number of allelic variations is five or greater, this method of subtracting the initial haplotype sequence from the composite genomic sequence of other individuals readily provides recognizably distinct haplotypes which are independent of each other. See, for example, the OMI¹ and GB haplotypes in FIG. 1, which differ from each other in each of seven sites of allelic variation.

When a haplotype determined in one individual is used to determine the haplotypes present in the composite genomic sequence of other individuals, the presence of that particular haplotype, and its companion haplotype as determined by subtraction from a composite genomic sequence, should be confirmed. Such confirmation of the occurrence of a given haplotype in the population can be carried out, for example, by 1) sequencing cDNA samples, as described in this section, from such other heterozygous individuals; or 2) identifying individuals homozygous for the haplotype either among the initial set of sequenced chromosomal samples or by additional confirmatory sequencing of chromosomal samples as described below.

If an initial haplotype is not represented in any heterozygous composite genomic sequences obtained, one or more additional haplotypes should be obtained from such a heterozygous individual or from different individuals screened as above.

cDNA sequences for determining the initial haplotype can be obtained using standard techniques well known in the art. First, mRNA is isolated from an individual, for example, from blood or skin cells. The mRNA is initially reversed-transcribed into double stranded cDNA and then amplified according to the well known technique of RT-PCR (see, for example, U.S. Pat. No. 5,561,058 by Gelfand et al.).

The resulting cDNA, whose sequence represents a single haplotype, can be sequenced according to the methods above.

Determining the Relative Frequencies of the Haplotype

After all haplotypes have been identified in the study population, their relative frequencies are determined. For example, if five chromosomes out of a total of ten chromosomes are of one haplotype, then its frequency is 50%. Subsequently, each haplotype is ranked in order from the most frequent to the least frequent to yield the functional allele profile.

Confirmatory Analysis of Additional Samples

As described above, initial sequence analysis is performed on a small group of individuals, most preferably five individuals, screened according to the methods described above.

After identifying the haplotypes and determining their relative frequencies among the initial set of alleles analyzed, it may be desirable, in accordance with the invention, to perform follow-up, confirmatory sequencing on additional individuals who are also screened according to the methods described above. Confirmatory sequencing can be carried out as above.

The haplotypes found occurring in the population are used as references to interpret the haplotypes present in any

A000554

US 6,951,721 B2

15

heterozygous individuals encountered during the confirmatory sequencing analysis of additional individuals.

By sequencing such additional samples, additional data points can be added to the functional allele profile to provide more precise frequencies of occurrence of each allele in the population. Furthermore, additional samples may contain a new functional allele with a new haplotype. This is particularly likely to be found for uncommon (<10%) or rare (<1%) haplotypes.

Furthermore, confirmatory sequence analysis ensures that the haplotypes determined by subtracting an initial haplotype from a composite heterozygous sequence is indeed represented in the population. Such techniques may also be used when multiple common haplotypes exist for the gene and it is uncertain which to use for subtraction.

When no sequence variation is found in the initial set of chromosomes, this indicates that the polymorphism rate of the gene of interest is uncommon (e.g., polymorphisms occur in <10% of the alleles in the population studied). In such situations, identification of uncommon alleles and determination of their frequencies requires a confirmatory sequence analysis of samples from additional individuals. This method was used to detect such an uncommon polymorphism in exon 8 of the MLH1 gene, in Example 2 below.

Such confirmatory sequencing analysis also resulted in the identification and determination of relative frequency of occurrence of polymorphisms in intronic sequences, bordering exonic regions, of both the MSH2 and MLH1 genes, as detailed in Examples 1 and 2, respectively, below. The invention is illustrated by way of the Examples below.

16

EXAMPLE 1

Determining the Functional Allele Profile for MSH2

Approximately 150 volunteers are screened in order to identify individuals with no cancer history in their immediate family (i.e. first and second degree relatives). Each person is asked to fill out the hereditary cancer prescreening questionnaire shown in Table 1, above. A first degree relative is a parent, sibling, or offspring. A second degree relative is an aunt, uncle, grandparent, grandchild, niece, nephew, or half-sibling. Among those individuals who answered "no" to all questions, five individuals are randomly chosen for end-to-end sequencing of their MSH2 gene.

Genomic DNA (100 nanograms) is extracted from white blood cells of five individuals designated as low risk of being carriers of mutations in the MSH2 gene from analysis of their answers to the questionnaire set forth in Table 1 above. The MSH2 coding region in each of the five samples is sequenced end-to-end by amplifying each exon individually. Each sample is amplified in a final volume of 25 microliters containing 1 microliter (100 nanograms) genomic DNA, 2.5 microliters 10×PCR buffer (100 mM Tris, pH 8.3, 500 mM KCl, 1.2 mM MgCl₂), 2.5 microliters 10×dNTP mix (2 mM each nucleotide), 2.5 microliters forward primer, 2.5 microliters reverse primer, and 1 microliter Taq polymerase (5 units), and 13 microliters of water.

The primers in Table 2, below, are used to carry out amplification of the various sections of the MSH2 gene samples. The primers are synthesized on an DNA/RNA Synthesizer Model 394@.

TABLE 2

MSH2 PRIMER SEQUENCES		
Ex on Primer	Sequence	
1 MSH1F-1	5'-CGC GTC TGC TTA TGA TTG G-3'	(SEQ ID NO: 1)
MSH1R-1	5'-TCT CTG AGG CGG GAA AGG-3'	(SEQ ID NO: 2)
2 MSH2-2F-2-INSIDE	5'-TTT TTT TTT TTT TAA GGA GC-3'	(SEQ ID NO: 3)
MSH2-2R-FULL	5'-CAC ATT TTT ATT TTT CTA CTC-3'	(SEQ ID NO: 4)
3 MSH3F	5'-GCT TAT AAA ATT TTA AAG TAT GTT C-3'	(SEQ ID NO: 5)
MSH3R-2	5'-CTG GAA TCT CCT CTA TCA C-3'	(SEQ ID NO: 6)
4 MSH4F	5'-TTC ATT TTT GCT TTT CTT ATT CC-3'	(SEQ ID NO: 7)
MSH4R	5'-ATA TGA CAG AAA TAT CCT TC-3'	(SEQ ID NO: 8)
5 MSH2-5F-1	5'-CAG TGG TAT AGA AAT CTT CGA-3'	(SEQ ID NO: 9)
MSH2-5R-2-INSIDE	5'-TTT TTT TTT TTT TTA CCT GA-3'	(SEQ ID NO: 10)
6 MSH6F-1	5'-ACT AAT GAG CTT GCC ATT CT-3'	(SEQ ID NO: 11)
MSH6R-1	5'-TGG GTA ACT GCA GGT TAC A-3'	(SEQ ID NO: 12)
7 MSH7F	5'-GAC TTA CGT GCT TAG TTG-3'	(SEQ ID NO: 13)
MSH7R	5'-AGT ATA TAT TGT ATG AGT TGA AGG-3'	(SEQ ID NO: 14)
8 MSH8F	5'-GAT TTG TAT TCT GTA AAA TGA GAT C-3'	(SEQ ID NO: 15)
MSH8R	5'-GGC CTT TGC TTT TTA AAA ATA AC-3'	(SEQ ID NO: 16)
9 MSH9F	5'-GTC TTT ACC CAT TAT TTA TAG G-3'	(SEQ ID NO: 17)
MSH9R	5'-GTA TAG ACA AAA GAA TTA TTC C-3'	(SEQ ID NO: 18)
10 MSH10F	5'-GGT AGT AGG TAT TTA TGG AAT AC-3'	(SEQ ID NO: 19)
MSH10R	5'-CAT GTT AGA GCA TTT AGG G-3'	(SEQ ID NO: 20)

A000555

US 6,951,721 B2

17

18

TABLE 2-continued

MSH2 PRIMER SEQUENCES		
Ex on Primer	Sequence	
11 MSH11F	5'-CAC ATT GCT TCT AGT ACA C-3'	(SEQ ID NO: 21)
MSH11R	5'-CCA GGT GAC ATT CAG AAC-3'	(SEQ ID NO: 22)
12 MSH12F	5'-ATT CAG TAT TCC TGT GTA C-3'	(SEQ ID NO: 23)
MSH12R	5'-CGT TAC CCC CAC AAA GC-3'	(SEQ ID NO: 24)
13 MSH13F-1	5'-ATG CTA TGT CAG TGT AAA CC-3'	(SEQ ID NO: 25)
MSH13R-1	5'-CCA CAG GAA AAC AAC TAT TA-3'	(SEQ ID NO: 26)
14 MSH14F	5'-TAC CAC ATT TTA TGT GAT GG-3'	(SEQ ID NO: 27)
MSH14R	5'-GGG GTA GTA AGT TTC CC-3'	(SEQ ID NO: 28)
15 MSH15F	5'-CTC TTC TCA TGC TGT CCC-3'	(SEQ ID NO: 29)
MSH15R	5'-ATA GAG AAG CTA AGT TAA AC-3'	(SEQ ID NO: 30)
16 MSH16F	5'-TAA TTA CTC ATG GGA CAT TC-3'	(SEQ ID NO: 31)
MSH16R-1	5'-GGC ACT GAC AGT TAA CAC TA-3'	(SEQ ID NO: 32)
NOTE: These MSH2 primers are M-13 tailed:		
M13 tail for	5'-TGT AAA ACG ACG GCC AGT-3' (SEQ ID NO: 33)	
F:	added to 5' end of primer above	
M13 tail for	5'-CAG GAA ACA GCT ATG ACC-3' (SEQ ID NO: 34)	
R:	added to 5' end of primer above	

Thirty-five cycles are performed, each consisting of dena-
turing (95° C.; 30 seconds), annealing (55° C.; 1 minute),
and extension (72° C.; 90 seconds), except during the first
cycle in which the denaturing time was increased to 5
minutes, and during the last cycle in which the extension
time was increased to 5 minutes.

PCR products are purified using Qia-quick® PCR puri-
fication kits (Qiagen®, cat#28104; Chatsworth, Calif.).
Yield and purity of the PCR product determined spectro-
photometrically at OD₂₆₀ on a Beckman DU 650 spectro-
photometer.

All exons of the MSH2 gene are subjected to direct
dideoxy sequence analysis by asymmetric amplification
using the polymerase chain reaction (PCR) to generate a
single stranded product amplified from this DNA sample.
Shuldiner, et al., Handbook of Techniques in Endocrine
Research, p. 457-486, DePablo, F., Scanes, C., eds., Aca-
demic Press, Inc., 1993. Fluorescent dye is attached to PCR
products for automated sequencing using the Taq Dye Ter-
minator Kit (Perkin-Elmer® cat#401628). DNA sequencing
is performed in both forward and reverse directions on an
Applied Biosystems, Inc. (ABI) Foster City, Calif., auto-
mated sequencer (Model 377). The software used for analy-
sis of the resulting data is "Sequence Navigator®" pur-
chased through ABI.

Results

No differences in nucleotide sequence are observed
among the coding exons of the five normal individuals (10
chromosomes), nor between these 10 chromosomal
sequences and the sequence published in GenBank
(Accession No. U03911) for MSH2. Thus, all ten individu-
als are homozygous for the same allele. An additional
sixty-two normal individuals are sequenced end-to-end to
confirm this result. Once again no sequence variation is
found within the exons. However, minor variation in three
single nucleotide polymorphisms are found in non-coding
intronic sequences (IVS9-9; IVS10+6; IVS 10+12). The
results are summarized in Table 3, below.

TABLE 3

MSH2 HAPLOTYPES					
<u>Allelic Variations</u>					
35	Haplotype	IVS9-9	IVS10 + 6	IVS10 + 12	Number of Chromosomes
	GenBank sequence (U03911)	T	T	A	98 (73%)
40	Variant #1	A	C	G	28 (21%)
	Variant #2	A	C	A*	6 (4.5%)
	Variant #3	T	C**	A	2 (1.5%)

*Variant #2 is an uncommon derivative chromosome of variant #1

**Variant #3 is a rarer derivative chromosome of GenBank cDNA

Since the exonic coding sequence is maintained on all 4
haplotypes, such non-coding sequence variation did not
result in any new "normal" coding consensus sequence of
the MSH2 gene.

These results demonstrate that the sequence in the Gen-
Bank Repository is the "consensus normal DNA sequence"
that should be used for comparison in all clinical applica-
tions to determine an individual with a hereditary suscepti-
bility to HNPCC. In addition, these results indicate that
normal MSH2 protein function, i.e., mismatch repair
function, is under a large degree of selective pressure to
maintain viability in the human population. Very little if any
variation in the activity of the MSH2 protein's mismatch
repair function is tolerated, as reflected by the extraordi-
narily high degree of conservation of the normal sequence.

EXAMPLE 2

Determining the Functional Allele Profile for MLH1

All procedures (e.g., selection of five individuals at low
risk of being carriers for MLH1 mutations, isolation of
genomic DNA, amplification of exons, sequencing of ampli-
fied exons, and analysis of sequence data) are carried out as
described in Example 1, above, except that the amplification
is carried out using primers specific to the MLH1 exons as
set forth in Table 4, below.

A000556

US 6,951,721 B2

19

20

TABLE 4

MLH1 PRIMER SEQUENCES		
Ex- on Primer	Sequence	
1 MLHAF	5'-AGG CAC TGA GGT GAT TGG C-3'	(SEQ ID NO: 35)
MLHAR	5'-TCG TAG CCC TTA AGT GAG C-3'	(SEQ ID NO: 36)
2 MLHBF-2	5'-TGA GGC ACT ATT GTT TGT ATT T-3'	(SEQ ID NO: 37)
MLHBR-2	5'-TGT TGG TGT TGA ATT TTT CAG T-3'	(SEQ ID NO: 38)
3 MLHCF	5'-AGA GAT TTG GAA AAT GAG TAA C-3'	(SEQ ID NO: 39)
MLHCR	5'-ACA ATG TCA TCA CAG GAG G-3'	(SEQ ID NO: 40)
4 MLHDF-1	5'-TGA GGT GAG AGT GGG TGA-3'	(SEQ ID NO: 41)
MLHGR	5'-GAT TAC TCT GAG ACC TAG GC-3'	(SEQ ID NO: 42)
5 MLHEF	5'-GAT TTT GTG TTT TCC CCT TGG G-3'	(SEQ ID NO: 43)
MLHER	5'-CAA ACA AAG CTT CAA CAA TTT AC-3'	(SEQ ID NO: 44)
6 MLHFF	5'-GGG TTT TAT TTT CAA GTA CTT GTA TG-3'	(SEQ ID NO: 45)
MLHFR	5'-GCT CAG CAA CTG TTC AAT GTA TGA GC-3'	(SEQ ID NO: 46)
7 MLHGF	5'-CTA-GTG TGT GTT TTT GGC-3'	(SEQ ID NO: 47)
MLHGR	5'-CAT AAC CTT ATC TCC ACC-3'	(SEQ ID NO: 48)
8 MLHHF	5'-CTC AGC CAT GAG ACA ATA AAT CC-3'	(SEQ ID NO: 49)
MLHHR	5'-GGT TCC CAA ATA ATG TGA TGG-3'	(SEQ ID NO: 50)
9 MLHIF-1	5'-GTT TAT GGG NAG GAA CCT TGT-3'	(SEQ ID NO: 51)
MLHIR-1	5'-TGG TCC CAT AAA ATT CCC TGT-3'	(SEQ ID NO: 52)
10 MLHJF	5'-CAT GAC TTT GTG TGA ATG TAG ACC-3'	(SEQ ID NO: 53)
MLHJR	5'-GAG GAG AGC CTG ATA GAA CAT CTG-3'	(SEQ ID NO: 54)
11 MLHKF	5'-GGG CTT TTT CTC CCC CTC CC-3'	(SEQ ID NO: 55)
MLHKR	5'-AAA ATC TGG GCT CTC AC-3'	(SEQ ID NO: 56)
12 MLH1-LAF-2-INSIDE	5'-TTT AAT ACA GAC TTT GCT AC-3'	(SEQ ID NO: 57)
MLH1-LBR	5'-GAA AAG CCA AAG TTA GAA GG-3'	(SEQ ID NO: 58)
13 MLHMF	5'-TGC AAC CCA CCA AAT TTG GC-3'	(SEQ ID NO: 59)
MLHMR	5'-CTT TCT CCA TTT CCA AAA CC-3'	(SEQ ID NO: 60)
14 MLHNF	5'-TGG TGT CTC TAG TTC TGG-3'	(SEQ ID NO: 61)
MLHNR	5'-CAT TGT TGT AGT AGC TCT GC-3'	(SEQ ID NO: 62)
15 MLHOF-2*	5'-GCA GAA CTA TGT CTG TCT CAT-3'	(SEQ ID NO: 63)
MLHOR	5'-CGG TCA GTT GAA ATG TCA G-3'	(SEQ ID NO: 64)
16 MLHPF	5'-CAT TTG GAT CCG TTA NAG C-3'	(SEQ ID NO: 65)
MLHPR	5'-CAC CCG GCT GGA AAT TTT ATT TG-3'	(SEQ ID NO: 66)
17 MLHQF	5'-GGA AAG GCA CTG GAG AAA TGG G-3'	(SEQ ID NO: 67)
MLHQR	5'-CCC TCC AGC ACA CAT GCA TGT ACC G-3'	(SEQ ID NO: 68)
18 MLHRF	5'-TAA GTA GTC TGT GAT CTC CG-3'	(SEQ ID NO: 69)
MLHRR	5'-ATG TAT GAG GTC GTG TCC-3'	(SEQ ID NO: 70)
19 MLHSF	5'-GAC ACC AGT GTA TGT TGG-3'	(SEQ ID NO: 71)
MLHSR*	5'-GAG AAA GAA GAA CAC ATC CC-3'	(SEQ ID NO: 72)

NOTE: MLH1 primers are M-13 tailed,

*EXCEPT for MLH1 primers MLHOF-2, MLHOR & MLHSR:

M13 tail for 5'-TGT AAA ACG ACG GCC AGT-3'

F: added to 5' end of primer above

M13 tail for 5'-CAG GAA ACA GCT ATG ACC-3'

R: added to 5' end of primer above

A000557

US 6,951,721 B2

21

Results

No differences are observed among the coding exons of the five normal individuals (10 chromosomes), nor between these 10 chromosomal sequences and the sequence published in GenBank (Accession No. U40978) for the MLH1 gene. In order to confirm these findings confirmatory sequencing is performed on an additional 62 samples. Among these sixty-two samples, variations are identified in only two positions as summarized in Table 5, below.

TABLE 5

Haplotype	MLH1 Haplotypes Allelic Variation		Number of Chromosomes
	EXON 8 codon 219	IVS14 - 19	
GenBank Sequence (040978)	A	A	114 (92.5%)
Variant #1	A	G	5 (3.7%)
Variant #2	G	G	4 (3.1%)
Variant #3	G	A	1 (0.7%)
Total 134 (100%)			

One sequence variation is within exon 8 wherein a single nucleotide change from A to G in the first position of codon 219 (ATC→GTC) changes the amino acid from Ile to Val. This sequence variation occurs approximately 3.7% of the time in this population. The second sequence variation is deep within an intron (IVS14-19) and can be found to be independently segregating with the exon 8 polymorphisms. While there were two "normal" exonic haplotypes identified

22

in MLH1 (A versus G at codon 219), the most commonly found haplotype (i.e. consensus normal DNA sequence) having an A at the first position of codon 219 is the sequence currently in the GenBank database which should be used as the standard for clinical comparisons.

In addition, this analysis demonstrated that there is less selective pressure on the MLH1 gene (since codon 219 can have two forms) than on the MSH2 gene where no exonic sequence variation was tolerated. Given that these two genes are both mismatch repair genes, this observation indicates that the degree of redundancy of function (i.e., level of hierarchy between these proteins) is MSH2 as the primary system with MLH1 only as secondary or backup when MSH2 is dysfunctional (i.e., mutant). While empiric data from other studies proposed such a relationship, only determining the actual functional allele profiles for these two genes provides an accurate understanding of the basis of previous observations from population studies.

EXAMPLE 3

Determining the Functional Allele Profile for
BRCA1

All procedures (e.g., selection of five individuals at low risk of being carriers for BRCA1 mutations, isolation of genomic DNA, amplification of exons, and sequencing of amplified exons, and analysis of sequence data) are carried out as described in Example 1, above, except that the amplification is carried out using primers specific to the BRCA1 exons as set forth in Table 6, below.

TABLE 6

BRCA1 PRIMERS FOR SEQUENCING TEMPLATES				
Exon	Primer	SEQUENCE	Mg ⁺⁺	SIZE
2	2F	5'GAAGTTGTCATTTTATAAACCTTT-3' (SEQ ID NO: 73)	1.6	275
	2R	5'TGCTCTTTTCTCCCTAGTATGT-3' (SEQ ID NO: 74)		
3	3F	5'TCCTGACACAGCAGACATTA-3' (SEQ ID NO: 75)	1.4	375
	3R	5'TTGGATTTCGTTCTCACTTTA-3' (SEQ ID NO: 76)		
5	5F	5'CTCTTAAGGGCAGTTGTGAG-3' (SEQ ID NO: 77)	1.2	275
	5R	5'TTCTACTGTGGTTGCTTCC-3' (SEQ ID NO: 78)		
6	6/7F	5'CTTATTTTAGTGTCTTAAAGG-3' (SEQ ID NO: 79)	1.6	250
	6R	5'TTTCATGGACAGCACTTGAGTG-3' (SEQ ID NO: 80)		
7	7F	5'CACAACAAAGAGCATACATAGGG-3' (SEQ ID NO: 81)	1.6	275
	6/7R	5'TCGGGTTCACCTCTGTAGAAG-3' (SEQ ID NO: 82)		
8	8F1	5'TTCTCTTCAGGAGGAAAAGCA-3' (SEQ ID NO: 83)	1.2	270
	8R1	5'GCTGCCTACCACAAATACAAA-3' (SEQ ID NO: 84)		
9	9F	5'CCACAGTAGATGCTCAGTAA TA-3' (SEQ ID NO: 85)	1.2	250
	9R	5'TAGGAAAATACCAGCTTCATAGA-3' (SEQ ID NO: 86)		
10	10F	5'TGGTCAGCTTTCTGTAATCG-3' (SEQ ID NO: 87)	1.6	250
	10R	5'GTATCTACCCACTCTCTTCTTCAG-3' (SEQ ID NO: 88)		
11A	11AF	5'CCACCTCCAAGGTGTATCA-3' (SEQ ID NO: 89)	1.2	372
	11AR	5'TGTTATGTTGGCTCCTTGCT-3' (SEQ ID NO: 90)		
11B	11BF1	5'CACTAAAGACAGAATGAATCTA-3' (SEQ ID NO: 91)	1.2	400
	11BR1	5'GAAGAAGCAGAATATTCATCTA-3' (SEQ ID NO: 92)		
11C	11CF1	5'TGATGGGGAGTCTGAATCAA-3' (SEQ ID NO: 93)	1.2	400
	11CR1	5'TCTGCTTTCTTGATAAAATCCT-3' (SEQ ID NO: 94)		

A000558

US 6,951,721 B2

23

24

TABLE 6-continued

BRCA1 PRIMERS FOR SEQUENCING TEMPLATES				
Exon	Primer	SEQUENCE	Mg ⁺⁺	SIZE
11D	11DF1	5'AGCGTCCCTCACAATAAA-3' (SEQ ID NO: 95)	1.2	400
	11DR1	5'TCAAGCGCATGAATATGCCT-3' (SEQ ID NO: 96)		
11E	11EF	5'GTATAAGCAATATGGAAGTCGA-3' (SEQ ID NO: 97)	1.2	388
	11ER	5'TTAAGTTCACTGGTATTGAACA-3' (SEQ ID NO: 98)		
11F	11FF	5'GACAGCGATACTTTCCCGA-3' (SEQ ID NO: 99)	1.2	382
	11FR	5'TGGAACAACCATGAATTAGTC-3' (SEQ ID NO: 100)		
11G	11GF	5'GGAAGTTAGCACTCTAGGGA-3' (SEQ ID NO: 101)	1.2	423
	11GR	5'GCAGTGATATTAAGTGTCTGTA-3' (SEQ ID NO: 102)		
11H	11HF	5'TGGGTCTTAAAGAAACAAAGT-3' (SEQ ID NO: 103)	1.2	366
	11HR	5'TCAGGTGACATTGAATCTTCC-3' (SEQ ID NO: 104)		
11I	11IF	5'CCACTTTTCCCATCAAGTCA-3' (SEQ ID NO: 105)	1.2	377
	11IR	5'TCAGGATGCTTACAATTACTTC-3' (SEQ ID NO: 106)		
11J	11JF	5'CAAATTGAATGCTATGCTTAGA-3' (SEQ ID NO: 107)	1.2	377
	11JR	5'TCGGTAACCCGTAGCCAAAT-3' (SEQ ID NO: 108)		
11K	11KF	5'GCAAAGCGTCCAGAAAGGA-3' (SEQ ID NO: 109)	1.2	396
	11KR-1	5'TATTTGCGACTCAAGCTTCCAA-3' (SEQ ID NO: 110)		
11L	11LF-1	5'GTAATATTGGCAAAGGCATCT-3' (SEQ ID NO: 111)	1.2	360
	11LR	5'TAAATGTGCTCCCCAAAGCA-3' (SEQ ID NO: 112)		
12	12F	5'GTCTGCCAATGAGAAGAAA-3' (SEQ ID NO: 113)	1.2	300
	12R	5'TGTGAGCAAAACCTAAGAATGT-3' (SEQ ID NO: 114)		
13	13F	5'AATGGAAGCTTCTCAAAGTA-3' (SEQ ID NO: 115)	1.2	325
	13R	5'ATGTTGGAGCTAGGTCTTAC-3' (SEQ ID NO: 116)		
14	14F	5'CTAACCTGAATTATCACTATCA-3' (SEQ ID NO: 117)	1.2	310
	14R	5'GTGTATAAATGCCTGTATGCA-3' (SEQ ID NO: 118)		
15	15F	5'TGGCTGCCAGGAAGTATG-3' (SEQ ID NO: 119)	1.2	375
	15R	5'AACCAGAATATCTTTATGTAGGA-3' (SEQ ID NO: 120)		
16	16F	5'AATTCTTAACAGAGACCAGAAC-3' (SEQ ID NO: 121)	1.6	550
	16R	5'AAACTCTTCCAGAATGTTGT-3' (SEQ ID NO: 122)		
17	17F	5'GTGTAGAACGTGCAGGATTG-3' (SEQ ID NO: 123)	1.2	275
	17R	5'TCGCCTCATGTGTTTAA-3' (SEQ ID NO: 124)		
18	18F	5'GGCTCTTTAGCTTCTTAGGAC-3' (SEQ ID NO: 125)	1.2	350
	18R	5'GAGACCATTTTCCAGCATC-3' (SEQ ID NO: 126)		
19	19F	5'CTGTCATCTTCTCTGTGCTC-3' (SEQ ID NO: 127)	1.2	250
	19R	5'CATTGTTAAGGAAAGTGGTGC-3' (SEQ ID NO: 128)		
20	20F	5'ATATGACGTGTCTGCTCCAC-3' (SEQ ID NO: 129)	1.2	425
	20R	5'GGGAATCCAAATTACACAGC-3' (SEQ ID NO: 130)		
21	21F	5'AAGCTCTTCCTTTTGAAGTC-3' (SEQ ID NO: 131)	1.6	300
	21R	5'GTAGAGAAATAGATAGCCTCT-3' (SEQ ID NO: 132)		
22	22F	5'TCCCATGAGAGGTCTTGCT-3' (SEQ ID NO: 133)	1.6	300
	22R	5'GAGAAGACTTCTGAGGTAC-3' (SEQ ID NO: 134)		
23	23F-1	5'TGAAGTGACAGTTCCAGTAGT-3' (SEQ ID NO: 135)	1.2	250
	23R-1	5'CATTTTAGCCATTCAATCAACAA-3' (SEQ ID NO: 136)		
24	24F	5'ATGAATTGACACTAATCTCTGC-3' (SEQ ID NO: 137)	1.4	285
	24R	5'GTAGCCAGGACAGTAGAAGGA-3' (SEQ ID NO: 138)		

¹M13 tailed

Results

Differences in the nucleotide sequences of the five normal individuals are found in seven locations on the gene. The data show that for each of the samples, the BRCA1 gene is identical except in the region of seven single nucleotide

polymorphisms. The changes and their positions are summarized on Table 7, below, and are depicted in schematic form in FIG. 1. The alternative alleles containing polymorphic (non-mutation causing allelic variations) sites along the BRCA1 gene are represented in FIG. 1 as individual "hap-

A000559

US 6,951,721 B2

25

lotypes" of the BRCA1 gene. The BRCA1^(omi1) haplotype is shown in FIG. 1 and indicated with dark shading. The alternative allelic variations occurring at nucleotide positions 2201, 2430, 2731, 3232, 3667, 4427, and 4956 are shown. For comparison, the haplotype previously available in GenBank (as Accession No. U14680) is completely unshaded and designated "GB". As can be seen, the most common, "consensus" haplotype occurs in five separate chromosomes labeled with the OMI symbol (haplotypes 1–5 from left to right). Two additional haplotypes (BRCA1^(omi2), and BRCA1^(omi3)) are represented with mixed shaded and unshaded positions (numbers 7 and 9 from left to right). In total, 7 of the ten 10 haplotypes identified in the group of five individuals tested are not the haplotype available in GenBank.

The changes, their positions, and their frequencies among the five individuals (ten chromosomes) initially analyzed are summarized on Table 7, below.

TABLE 7

NORMAL PANEL TYPING									
AMINO ACID CHANGE	EXON	1	2	3	4	5	FREQUENCY		
SER(SER) (694)	11E	C/C	C/T	C/T	T/T	T/T	0.4	C	0.6 T
LEU(LEU) (771)	11F	T/T	C/T	C/T	C/C	C/C	0.4	T	0.6 C
PRO(LEU) (871)	11G	C/T	C/T	C/T	T/T	T/T	0.3	C	0.7 T
GLU(GLY) (1038)	11I	A/A	A/G	A/G	G/G	G/G	0.4	A	0.6 G
LYS(ARG) (1183)	11J	A/A	A/G	A/G	G/G	G/G	0.4	A	0.6 G
SER(SER) (1436)	13	T/T	T/T	T/C	C/C	C/C	0.5	T	0.5 C
SER(GLY) (1613)	16	A/A	A/G	A/G	G/G	G/G	0.4	A	0.6 G

Note that there is no requirement to sequence the additional normal individuals available, as has been done for MSH2 (Example 1, above) and MLH1 (Example 2, above) to more accurately determine the frequencies of uncommon polymorphisms. A common haplotype (the "consensus") is readily evident as different from the GenBank sequence (FIG. 1, "GB") in 50% of chromosomes and indeed is homozygous in two normal individuals.

Thus, the "consensus" sequence of the BRCA (omi¹) should be used as the only true standard for clinical diagnostic analysis in order to avoid misinterpreting polymorphisms as pathologic mutations.

In the alternative, one could compare the test sequence against all four of the BRCA1 functional haplotypes.

EXAMPLE 4

Pharmacogenetic Analysis of Sulfa Drug Sensitivity

The glucose-6-phosphate dehydrogenase gene is located on the X chromosome. Individuals with certain sequence variations in the G6PDH gene lead relatively normal lives unless they are exposed to certain chemicals found in fava beans, primaquine and sulfonamide antibiotics (sulfisoxazole, sulfamethoxazole, sulfathiazole, sulfacetamide, etc.). Upon administration of such compounds to the individual, severe reactions including hemolytic anemia occur in individuals having certain haplotype(s) of the G6PDH gene. These individuals are

26

generally of African and Mediterranean heritage. Because these sequence variations are otherwise of little importance, they have been called both polymorphisms and mutations in the literature. For the purposes of this application, they are called mutations to distinguish them from clear polymorphisms. Genetic analysis in chimpanzees and various human populations indicate that the probable natural "wild-type" is found in individuals sensitive to sulfonamide antibiotics. Beutler et al, *Blood* 74: 2550–2555 (1989).

A number of apparently inconsequential single nucleotide polymorphisms (SNPs) in the G6PDH gene are known including at intron 5 (PvuII site), nucleotides 202 (Nla III site), 376 (Fok I site), 1311 and 1116 (Pst I sites). These constitute and define the haplotype. Missense mutations occur at amino acids 32, 48, 58, 68, 106, 126, 131, 156, 163, 165, 181, 182, 188, 198, 213, 216, 227, 282, 285, 291, 317, 323, 335, 342, 353, 363, 385, 386, 387, 393, 394, 398, 410, 439, 447, 454, 459, 463 and amino acid 35 deleted. Many mutations are restricted to certain haplotypes. Thus, haplotype determination provides an indication of whether the individual is sensitive to the drugs listed above.

Experimental

Blood is drawn from 30 individuals of African-American heritage with urinary tract infections having bacteria sensitive to sulfa antibiotics and for whom treatment with trimethoprim-sulfamethiazole is otherwise deemed appropriate. 1 mg of genomic DNA from individuals is isolated from peripheral blood lymphocytes and amplified by PCR using the primers listed in Hirono et al, *Proc. Natl. Acad. Sci. USA* 85:3951–3954 (1988) and Beutler et al, *Human Genetics* 87:462–464 (1990) according to the methods in Example 1 above. Amplified fragments are divided into five aliquots and four of which are cleaved by a restriction enzyme, either PvuII, Nla III, Fok I or Pst I, according to the manufacturer's (Stratagene and New England Biolabs) instructions. The digests are electrophoresed in a 4% agarose gel (NuSieve, FMC) with 10 ml of ethidium bromide (10 mg/ml) and the number of bands counted under ultraviolet light. The number of bands indicates the presence or absence of restriction enzyme cleavage and presence of a particular nucleotide at the polymorphic site.

An oligonucleotide probe for determining the polymorphic site at nucleotide 1311 is listed in Beutler et al, *Human Genetics* 87:462–464 (1990). The fifth aliquot is immobilized on a membrane and an ASO (allele specific oligonucleotide) hybridization assay is performed according to the method of Example 5 below. The presence or absence of the label indicating hybridization is considered indicative of the presence of a particular nucleotide at the polymorphic site.

Individuals having a haplotype, particularly the polymorphism at nucleotide 1116, indicative of very low likelihood of a G6PDH mutation sensitive to sulfamethiazole are given 160 mg trimethoprim with 800 mg sulfamethiazole (SEPTA DS). Individuals having a haplotype or polymorphism indicative of a possible presence of a G6PDH mutation sensitive to sulfamethiazole are given a different antibiotic (varied with the patient) to which their infecting organism was susceptible.

Confirmatory sequencing of both alleles (60 chromosomes) of the coding region of the G6PDH gene is later performed by the techniques of Example 1 to determine

A000560

US 6,951,721 B2

27

the presence of a sensitizing mutation. The haplotype(s) associated with a mutation and those not associated with a mutation are recorded. A panel of oligonucleotides bound to a membrane or other solid phase such as a DNA chip distinguishing the haplotypes and/or the common mutations also is to become part of the present invention.

EXAMPLE 5

Pharmacogenetic Analysis of BRCA1, BRCA2, PTEN, BAP1, BARD1 and hRAD51 Haplotypes and the Use of Tamoxifen to Prevent Breast Cancer

While every step in carcinogenesis is not known, the BRCA1, BRCA2, PTEN, BAP1, BARD1 and hRAD51 proteins are either involved in breast, ovarian, prostate and other cancer susceptibility, in the metabolic pathway of or interact with such proteins. It was determined that the most common form of hereditary breast and ovarian cancer, the BRCA1 185delAG mutation, was found essentially exclusively in one haplotype, namely haplotype OMI1 as defined in Example 1, FIG. 1 and U.S. Pat. No. 5,654,155. As such it was applicants hypothesis that the haplotypes of other

28

breast cancers. Since some of these proteins actually bind to each other, different combinations of haplotypes may bind with different avidity to each other and operate slightly differently under certain circumstances. Likewise for proteins which act at separate reactions within the tumor-suppressing mechanisms.

Experimental

Blood samples are drawn from 47 women prescribed tamoxifen to prevent breast cancer or having had breast cancer to prevent reoccurrence of breast cancer. The DNA sequence for BRCA1 is determined in the regions of the single nucleotide polymorphic sites which constitute the haplotype use the primers according to U.S. Pat. No. 5,654,155. Those of BRCA2 are determined by using the primers of U.S. patent application Ser. No. 09/084,471 filed May 22, 1998 or using the primers:

TABLE 8

BRCA2 PRIMERS		
EXON	SEQUENCE	POLY-MORPHISM
10AF	5'GAATAATATAAATTATATGGCTTA-3' (SEQ ID NO: 139)	1093
10AF	5'CCTAGTCTTGCTAGTTCTT-3' (SEQ ID NO: 140)	1093
10BE	5'ARCTGAAGTGGAAACCAATGATAC-3' (SEQ ID NO: 141)	1593
10BR	5'ACGTGGCAAAGAATTCTCTGAAGTAA-3' (SEQ ID NO: 142)	1593
11BF	5'AAGAAGCAAATGTAATAAGGA-3' (SEQ ID NO: 143)	2457
11BR	5'CATTTAAGCACATACATCTTG-3' (SEQ ID NO: 144)	2457
11CF	5'TCTAGAGGCAAAGAATCATAC-3' (SEQ ID NO: 145)	2908
11CR	5'CAAGATTATTCCTTCATTAGC-3' (SEQ ID NO: 146)	2908
11DF	5'AACCAAAACACAAATCTAAGAG-3' (SEQ ID NO: 147)	3199
11DR	5'GTCATTTTATATGCTGCTTAC-3' (SEQ ID NO: 148)	3199
11EF	5'GGTTTTATATGGAGACACAGG-3' (SEQ ID NO: 149)	3624
11ER	5'GTATTTACAATTCAACACAAGC-3' (SEQ ID NO: 150)	3624
11FF	5'ATCACAGTTTTGGAGGTAGC-3' (SEQ ID NO: 151)	4035
11FR	5'CTGACTTCCTGATTCTTCTAA-3' (SEQ ID NO: 152)	4035
14F	5'ACCATGTAGCAAATGAGGGTCT-3' (SEQ ID NO: 153)	7470
14R	5'GCTTTTGTCTGTTTCCCTCAA-3' (SEQ ID NO: 154)	7470
22F	5'AACCACACCCTTAAGATGA-3' (SEQ ID NO: 155)	9079
22R	5'GCATAAGTAGTGGATTTTGC-3' (SEQ ID NO: 156)	9079

related and similar genes alone or in certain combinations provide an indication of association with breast and other cancers associated with these genes, e.g. ovarian, pancreatic, prostate, colon, etc.

The DNA sequences for haplotypes of PTEN are determined by using the published primers of Table 3, Liaw et al, *Nature Genetics*, 16(1): p. 64-67 (1997).

The primers for amplifying hRAD51 are:

5'GGGCCCCGATCCATGGCAATGCAGATGCAGC-3' (SEQ ID NO: 157) and

5'GGGCCCCAATGGATATCATTAGCTTTGGCATCTCCCACTCC-3' (SEQ ID NO: 158).

The various treatments and prophylactics useful against the disease are also believed to be related to the haplotypes. It is already known that certain mutant genes result in different presentations of cancers and different treatment. For example, BRCA1 mutations in the early part of the coding sequence generally form cancers at a younger age than mutations in the later part of the coding sequence. Likewise, breast cancer arising from BRCA2 mutations are typically more sensitive to radiation treatment than other

The primers for amplifying BAP1 are:

PRIMER	SEQUENCE
BAP1A1-F	5'CACGAGGCATGGCGCTGAGG-3' (SEQ ID NO: 159)
BAP1A-R	5'CCGGGCCTTGCTGTCCACT-3' (SEQ ID NO: 160)

US 6,951,721 B2

29

-continued

PRIMER	SEQUENCE
BAP1B-F	5'GTCTACCCCATTTGACCATGG-3' (SEQ ID NO: 161)
BAP1B-R	5'TCATCATCTGAGTACTGCTG-3' (SEQ ID NO: 162)
BAP1C-F	5'TGCAGGAGGAAGAAGACCTG-3' (SEQ ID NO: 163)
BAP1C-R	5'TCTGTCAGCGCCAGGGGACT-3' (SEQ ID NO: 164)
BAP1D-F	5'AGCACAGGCCTGCTGCACCT-3' (SEQ ID NO: 165)
BAP1D-R	5'GAAAAGGGGAAGTGGGGCAG-3' (SEQ ID NO: 166)

The primers for amplifying BAP1 for polymorphism detection in the 3' UTR are:

BAP1-PF 5'AGCCAGGCCCAACACAGCCCATGGCCTCT-3' (SEQ ID NO: 167)

BAP1-PR 5'CTTAGGAGAGTTTATTTCATTGATCCAG-3' (SEQ ID NO: 168)

The primers for amplifying BARD1 are:

5'AACAGTACAATGACTGGGCTC -3' (SEQ ID NO: 169) and

5'TCAGCGCTTCTGCACACAGT -3' (SEQ ID NO: 170)

In the cases of BARD1 and hRAD51, the PCR products are sequenced in entirety. All procedures (e.g., isolation of genomic DNA, amplification, sequencing, and analysis of sequence data) are carried out as described in Example 1. The method as described in Examples 1-3 is used to determine the common haplotypes in these genes.

Once standardized by sequencing, the amplified fragments of BRCA1, BRCA2, PTEN and BAP1, produced by PCR are assayed by hybridization to allele-specific oligonucleotides (ASO) which distinguish the polymorphic site directly. The ASO assay is performed as described in the following experiment.

Binding PCR Products to Nylon Membrane

The PCR products are denatured no more than 30 minutes prior to binding the PCR products to the nylon membrane. To denature the PCR products, the remaining PCR reaction (45 ml) and the appropriate positive control mutant gene amplification product are diluted to 200 ml final volume with PCR Diluent Solution (500 mM NaOH, 2.0 M NaCl, 25 mM EDTA) and mixed thoroughly. The mixture is heated to 95° C. for 5 minutes, and immediately placed on ice and held on ice until loaded onto dot blotter, as described below.

The PCR products are bound to 9 cm by 13 cm nylon ZETA PROBE BLOTting MEMBRANE (BIO-RAD, Hercules, Calif., catalog number 162-0153) using a BIO-RAD dot blotter apparatus. Forceps and gloves are used at all times throughout the ASO analysis to manipulate the membrane, with care taken never to touch the surface of the membrane with bare hands or latex gloves.

Pieces of 3MM filter paper [WHATMAN®, Clifton, N.J.] and nylon membrane are pre-wet in 10xSSC prepared fresh from 20xSSC buffer stock. The vacuum apparatus is rinsed thoroughly with dH₂O prior to assembly with the membrane. 100 ml of each denatured PCR product is added to the wells of the blotting apparatus. Each row of the blotting apparatus contains a set of reactions for a single exon to be

30

tested, including a placental DNA (negative) control, a synthetic oligonucleotide with the desired mutation or a PCR product from a known mutant sample (positive control), and three no template DNA controls.

After applying PCR products, the nylon filter is placed DNA side up on a piece of 3MM filter paper saturated with denaturing solution (1.5M NaCl, 0.5 M NaOH) for 5 minutes. The membrane is transferred to a piece of 3MM filter paper saturated with neutralizing solution (1M Tris-HCl, pH 8, 1.5 M NaCl) for 5 minutes. The neutralized membrane is then transferred to a dry 3MM filter DNA side up, and exposed to ultraviolet light (STRALINKER, STRATAGENE, La Jolla, Calif.) for exactly 45 seconds to fix the DNA to the membrane. This UV crosslinking should be performed within 30 min. of the denaturation/neutralization steps. The nylon membrane is then cut into strips such that each strip contains a single row of blots of one set of reactions for a single exon.

Hybridizing Labeled Oligonucleotides to the Nylon Membrane Prehybridization

The strip is prehybridized at 52° C. incubation using the HYBAID® (SAVANT INSTRUMENTS, INC., Holbrook, N.Y.) hybridization oven. 2xSSC (15 to 20 ml) is preheated to 52° C. in a water bath. For each nylon strip, a single piece of nylon mesh cut slightly larger than the nylon membrane strip (approximately 1"x5") is pre-wet with 2xSSC. Each single nylon membrane is removed from the prehybridization solution and placed on top of the nylon mesh. The membrane/mesh "sandwich" is then transferred onto a piece of Parafilm™. The membrane/mesh sandwich is rolled lengthwise and placed into an appropriate HYBAID® bottle, such that the rotary action of the HYBAID® apparatus caused the membrane to unroll. The bottle is capped and gently rolled to cause the membrane/mesh to unroll and to evenly distribute the 2xSSC, making sure that no air bubbles formed between the membrane and mesh or between the mesh and the side of the bottle. The 2xSSC is discarded and replaced with 5 ml TMAC Hybridization Solution, which contained 3 M TMAC (tetramethyl ammoniumchloride-SIGMA T-3411), 100 mM Na₃PO₄(pH 6.8), 1 mM EDTA, 5x Denhardt's (1% Ficoll, 1% polyvinylpyrrolidone, 1% BSA (fraction V)), 0.6% SDS, and 100 mg/ml Herring Sperm DNA. The filter strips were prehybridized at 52° C. with medium rotation (approx. 8.5 setting on the HYBAID® speed control) for at least one hour. Prehybridization can also be performed overnight.

Labeling Oligonucleotides

The DNA sequences of the oligonucleotide probes used to detect the BRCA1, BRCA2, PTEN, and BAP1 single nucleotide polymorphisms (SNPs) are as follows (for each polymorphism both options for the oligonucleotide are given below): The complements of these probes may also be used. Preliminary laboratory data indicates that probes with either greater specificity or sensitivity can be prepared by slightly varying the length and amount overlapping each side of the polymorphic region. It is expected that better probes will be prepared by routine experimentation.

A000562

US 6,951,721 B2

31

TABLE 9

BRCA1			
2201	C	5'ACATGACAGCGATACTT-3'	(SEQ ID NO: 171)
2201	T	5'ACATGACAGTGATACTT-3'	(SEQ ID NO: 172)
2430	T	5'AGTATTTCATTGGTACC-3'	(SEQ ID NO: 173)
2430	C	5'AGTATTTCAGTGGTACC-3'	(SEQ ID NO: 174)
2731	C	5'CATTGTCTCCGTTTTC-3'	(SEQ ID NO: 175)
2731	T	5'CATTGTCTCTGTTTTC-3'	(SEQ ID NO: 176)
3232	A	5'TTTTAAAGAGGCCAGC-3'	(SEQ ID NO: 177)
3232	G	5'TTTTAAAGAGGCCAGC-3'	(SEQ ID NO: 178)
3667	A	5'GCGTCCAGAAAGGAGAG-3'	(SEQ ID NO: 179)
3667	G	5'GCGTCCAGAGAGGAGAG-3'	(SEQ ID NO: 180)
4427	T	5'AAGTGACTCTTCTGCCC-3'	(SEQ ID NO: 181)
4427	C	5'AAGTGACTCTCTGCCC-3'	(SEQ ID NO: 182)
4956	A	5'TGTGCCCAGAGTCCAGC-3'	(SEQ ID NO: 183)
4956	G	5'TGTGCCCAGGTCCAGC-3'	(SEQ ID NO: 184)
1186	A	5'GGAATAAGCAGAACTG-3'	(SEQ ID NO: 185)
1186	G	5'GGAATAAGCGAACTG-3'	(SEQ ID NO: 186)
2196	G	5'AAAAGACATGACAGCGA-3'	(SEQ ID NO: 187)
2196	A	5'AAAAGACATAACAGCGA-3'	(SEQ ID NO: 188)
3238	G	5'AAGAAGCCAGCTCAAGC-3'	(SEQ ID NO: 189)
3238	A	5'AAGAAGCCAAGCTCAAGC-3'	(SEQ ID NO: 190)
2202	G	5'CATGACAGTGATACTTT-3'	(SEQ ID NO: 191)
2202	A	5'CATGACAGTAATACTTT-3'	(SEQ ID NO: 192)

TABLE 10

BRCA2			
PROBE		SEQUENCE	
1093	A	5'TAGGACATGGCATTGA-3'	(SEQ ID NO: 193)
1093	C	5'TAGGACATGGGCATTGA-3'	(SEQ ID NO: 194)
1342	A	5'CTTCTGATTGTGCTACATT-3'	(SEQ ID NO: 195)
1342	C	5'CTTCTGATGTGCTACATT-3'	(SEQ ID NO: 196)
1593	A	5'GGCTTCTCTGATTTTGGT-3'	(SEQ ID NO: 197)
1593	G	5'GGCTTCTCGGATTTTGGT-3'	(SEQ ID NO: 198)
2457	T	5'TTTTGAATATTGTACTGG-3'	(SEQ ID NO: 199)
2457	C	5'TTTTGAATGTTGTACTGG-3'	(SEQ ID NO: 200)
2908	G	5'ATTAGCTACTTGAAGAC-3'	(SEQ ID NO: 201)
2908	A	5'ATTAGCTATTTGAAGAC-3'	(SEQ ID NO: 202)
3199	A	5'CCATTGTTCATGTAATC-3'	(SEQ ID NO: 203)
3199	G	5'CCATTGTTCATGTAATC-3'	(SEQ ID NO: 204)
3624	A	5'TAGCTTGGTTTCTTAAAC-3'	(SEQ ID NO: 205)
3624	G	5'TAGCTTGGTTTCTTAAAC-3'	(SEQ ID NO: 206)
4035	T	5'ATTGAAACACAGAATCA-3'	(SEQ ID NO: 207)
4035	C	5'ATTGAAACGACAGAATCA-3'	(SEQ ID NO: 208)
7470	A	5'TGAAAATGTGATTAGTT-3'	(SEQ ID NO: 209)
7470	G	5'TGAAAATGCGATTAGTT-3'	(SEQ ID NO: 210)
9079	G	5'TTCCATTGGCTTCTTAAT-3'	(SEQ ID NO: 211)
9079	A	5'TTCCATTGGTCTTCTTAAT-3'	(SEQ ID NO: 212)

32

TABLE 11

PTEN			
132	C	5'CTTGAAGCGTATACAGG-3'	(SEQ ID NO: 213)
132	T	5'CTTGAAGGTGTATACAGG-3'	(SEQ ID NO: 214)

TABLE 12

BAP1			
+1102	5'	ATGGCCTCTACAGATGGC-3'	(SEQ ID NO: 215)
+1102	5'	ATGGCCTCTCCAGATGGC-3'	(SEQ ID NO: 216)
+1102	5'	ATGGCCTCTCCAGATGGC-3'	(SEQ ID NO: 217)
+1102	5'	ATGGCCTCTCCAGATGGC-3'	(SEQ ID NO: 218)
+1116	5'	CAGATGGCTTTGAAAAGG-3'	(SEQ ID NO: 219)
+1116	5'	CAGATGGCTTTGAAAAGG-3'	(SEQ ID NO: 220)
+1116	5'	CAGATGGCTTTGAAAAGG-3'	(SEQ ID NO: 221)
+1116	5'	CAGATGGCTTTGAAAAGG-3'	(SEQ ID NO: 222)
+1131	5'	GATCCAAACAGGCCCTTT-3'	(SEQ ID NO: 223)
+1131	5'	GATCCAAACAGGCCCTTT-3'	(SEQ ID NO: 224)
+1131	5'	GATCCAAACAGGCCCTTT-3'	(SEQ ID NO: 225)
+1131	5'	GATCCAAACAGGCCCTTT-3'	(SEQ ID NO: 226)
+1233	5'	CCCTGTAAACTGGATCA-3'	(SEQ ID NO: 227)
+1233	5'	CCCTGTAAACTGGATCA-3'	(SEQ ID NO: 228)
+1233	5'	CCCTGTAAACTGGATCA-3'	(SEQ ID NO: 229)
+1233	5'	CCCTGTAAACTGGATCA-3'	(SEQ ID NO: 230)

Each labeling reaction contains 2- μ l 5 \times Kinase buffer (or 1 μ l of 10 \times Kinase buffer), 5 μ l gamma-ATP 32 P (not more than one week old), 1 μ l T4 polynucleotide kinase, 3 μ l oligonucleotide (20 μ M stock), sterile H₂O to 10 μ l final volume if necessary. The reactions are incubated at 37 $^{\circ}$ C. for 30 minutes, then at 65 $^{\circ}$ C. for 10 minutes to heat inactivate the kinase. The kinase reaction is diluted with an equal volume (10 μ l) of sterile dH₂O (distilled water).

The oligonucleotides are purified on STE MICRO SELECT-D, G-25 spin columns (catalog no. 5303-356769), according to the manufacturer's instructions. The 20 μ l synthetic oligonucleotide eluate is diluted with 80 μ l dH₂O (final volume=100 μ l). The amount of radioactivity in the oligonucleotide sample is determined by measuring the radioactive counts per minute (cpm). The total radioactivity must be at least 2 million cpm. For any samples containing less than 2 million total, the labeling reaction is repeated.

Hybridization with Oligonucleotides

Approximately 2-5 million counts of the labeled oligonucleotide probe is diluted into 5 ml of TMAC hybridization solution, containing 40 μ l of 20 μ M stock of unlabeled alternative polymorphism oligonucleotide. The probe mix is preheated to 52 $^{\circ}$ C. in the hybridization oven. The pre-hybridization solution is removed from each bottle and replaced with the probe mix. The filter is hybridized for 1 hour at 52 $^{\circ}$ C. with moderate agitation. Following hybridization, the probe mix is decanted into a storage tube and stored at -20 $^{\circ}$ C. The filter is rinsed by adding approximately 20 ml of 2 \times SSC+0.1% SDS at room temperature and rolling the capped bottle gently for approximately 30 seconds and pouring off the rinse. The filter is then washed with 2 \times SSC+0.1% SDS at room temperature for 20 to 30 minutes, with shaking.

The membrane is removed from the wash and placed on a dry piece of 3MM WHATMAN filter paper then wrapped in one layer of plastic wrap, placed on the autoradiography film, and exposed for about five hours depending upon a

A000563

US 6,951,721 B2

33

survey meter indicating the level of radioactivity. The film is developed in an automatic Film processor.

Control Hybridization with Normal Oligonucleotides

The purpose of this step is to ensure that the PCR products are transferred efficiently to the nylon membrane.

Following hybridization with the bound oligonucleotide, as described above, each nylon membrane is washed in 2×SSC, 0.1% SDS for 20 minutes at 65° C. to melt off the bound oligonucleotide probes. The nylon strips are then prehybridized together in 40 ml of TMAC hybridization solution for at least 1 hour at 52° C. in a shaking water bath. 2–5 million counts of each of the normal labeled oligonucleotide probes plus 40 μ l of 20 μ M stock of unlabeled normal oligonucleotide are added directly to the container containing the nylon membranes and the prehybridization solution. The filter and probes are hybridized at 52° C. with shaking for at least 1 hour. Hybridization can be performed overnight, if necessary. The hybridization solution is poured off, and the nylon membrane is rinsed in 2×SSC, 0.1% SDS for 1 minute with gentle swirling by hand. The rinse is poured off and the membrane is washed in 2×SSC, 0.1% SDS at room temperature for 20 minutes with shaking.

The nylon membrane is removed and placed on a dry piece of 3MM WHATMAN filter paper. The nylon membrane is then wrapped in one layer of plastic wrap and placed on autoradiography film. The exposure is for at least 1 hour.

For each sample, adequate transfer to the membrane is indicated by a strong autoradiographic hybridization signal. For each sample, an absent or weak signal when hybridized with its normal oligonucleotide, indicates an unsuccessful transfer of PCR product, and it is a false negative. The ASO analysis must be repeated for any sample that did not successfully transfer to the nylon membrane.

The pattern of hybridization using the probes from the panel according to Tables 9–12 determine the haplotype of the patient sample when compared to the known haplotypes.

The degree of breast, ovarian and other cancer prevention with and without tamoxifen and the degree of prevention of reoccurrence of breast and ovarian cancer with and without tamoxifen are compared for patients grouped by BRCA1, BRCA2, PTEN, BAP1, BARD1, hRAD51 haplotype separately and in all possible combinations using various proprietary data mining techniques similar to the Recognizer™ methodology described in U.S. Pat. No. 5,642,936. Appropriate recommendations regarding the use of tamoxifen for patients of different haplotypes are then be made for patients with and without a history of breast or ovarian cancer.

While this example is a retrospective study and thus unacceptable for proof of efficacy for the U.S. Food and Drug Administration, prospective studies are also part of the present invention. In a prospective study, the test individuals have their haplotypes determined for each pertinent gene prior to determining whether or not they will be accepted for the drug trial or initiate tamoxifen therapy.

EXAMPLE 6

Pharmacogenetic Analysis of a p53 Polymorphism and the Appropriateness of the Human Papilloma Virus Vaccine

Human papilloma virus (HPV) currently infects up to 40 million Americans with at least one of about 80 different strains. Many strains of the virus cause venereal warts, vulval, penile and perianal cancers. One strain in particular, HPV-16, is believed to be responsible for about half of all cases

34

of cervical cancer. Three other strains are responsible for another 35% of all cervical cancer cases with HPV-18 causing malignant tumors while HPV-6 and HPV-11 usually forming benign lesions. HPV vaccines are made by MedImmune, Inc. (Gaithersburg, Md.) and Merck & Co. Clinical trials have already begun.

While applicant does not wish to be bound by any theory, it is believed that HPV may induce cancer by interacting with p53 in a manner which inhibits the action of p53 to prevent runaway cell growth. It has been known that HPV protein E6 inactivates only p53 proteins from some individuals and not other individuals. Medcalf et al, *Oncogene*, 8: 2847–2851 (1993). Therefore, determining the haplotype(s) of the p53 gene is believed to indicate who is susceptible to cervical cancer induced by HPV and is therefore a candidate for a HPV vaccine.

Previous commercial p53 gene testing of patient samples performed by Oncormed, Inc. (the owner of this application) involved various sequencing techniques and functional assays for prognostic testing on various tumor samples and susceptibility testing of genomic samples in patients with an inherited mutant p53 gene (Li-Fraumeni Syndrome). While apparent single nucleotide polymorphisms were noticed, such results were not reported as the samples are suspected to contain p53 mutations and do not originate from healthy individuals without a genetic history indicating inheritance of two functional p53 alleles.

Only polymorphisms in the coding region are analyzed because women having cervical cancers are believed to have a p53 protein which is “in-activatable” because the coding sequence for p53 is usually not mutated in cervical cancers. Vogelstein et al, *Cell*, 70: 523–526 (1992). Thus, the haplotypes were determined based on the single nucleotide polymorphisms at codon 21 (which may be either GAC or GAT), codon 36 (which may be either CCG or CCA), codon 47 (which may be either CCG or TCG), codon 72 (which may be either CGC or CCC) and codon 213 (which may be either CGA or CGG).

Experimental Protocol

Blood samples are from 53 healthy individuals having a history of venereal warts or at risk from exposure to HPV. Exposure is defined as an individual having regular sexual contact with an infected individual without a barrier preventing transmission of HPV. These individuals have either stage I (normal) or stage II (inflammation) PAP smears. Some of the individuals had been previously treated for venereal warts with one or more of the following treatments: podophyllin, trichloroacetic acid, cryosurgery, cauterization or interferon. Also, blood samples are from 12 patients with a history of cervical cancer as defined by a stage IV (carcinoma in-situ) or greater PAP smear result. Note that individuals having a stage III PAP smear (dysplasia) are not included in this study. White blood cells are collected and genomic DNA is extracted from the white blood cells according to well-known methods (Sambrook, et al., *Molecular Cloning, A Laboratory Manual*, 2nd Ed., 1989, Cold Spring Harbor Laboratory Press, at 9.16–9.19).

PCR Amplification for Sequencing

The genomic DNA is used as a template to amplify a DNA fragment encompassing the site of the mutation to be tested. The 25 ml PCR reaction contains the following components: 1 ml template (100 ng/ml) DNA, 2.5 ml 10×PCR Buffer

A000564

US 6,951,721 B2

35

(PERKIN-ELMER), 1.5 ml dNTP (2 mM each dATP, dCTP, dGTP, dTTP), 1.5 ml Forward Primer (10 mM), 1.5 ml Reverse Primer (10 mM), 0.5 ml (2.5 U total) AMPLITAQ GOLD™ TAQ DNA POLYMERASE or AMPLITAQ® TAQ DNA POLYMERASE (PERKIN-ELMER), 1.0 to 5.0 ml (25 mM) MgCl₂ (depending on the primer) and distilled water (dH₂O) up to 25 ml. All reagents for each exon except the genomic DNA can be combined in a master mix and aliquoted into the reaction tubes as a pooled mixture. The primers are listed below.

NAME	SEQUENCE		LENGTH	IN-TRON
2F	5'-TCATGCTGGATCCCACTTTCTCTCTTG-3'	(SEQ ID NO: 231)	28	31
2R	5'-GGTGGCCTGCCTTCCAATGGATCCACT-3'	(SEQ ID NO: 232)	28	3
3F	5'-AATTCATGGGACTGACTTTCTGCTCTTGTC-3'	(SEQ ID NO: 233)	30	6
3R	5'-TCCAGGTCCCAGCCCAACCTTGTCC-3'	(SEQ ID NO: 234)	26	4
4F	5'-GTCCTCTGACTGCTCTTTTCAACCATCTAC-3'	(SEQ ID NO: 235)	30	2
4R	5'-GGGATACGGCCAGGCATTGAAGTCTC-3'	(SEQ ID NO: 236)	26	29
5F	5'-CTTGTGCCCTGACTTTCAACTCTGTCTC-3'	(SEQ ID NO: 237)	28	16
5R	5'-TGGGCAACCAGCCCTGTCGCTCTCTCCA-3'	(SEQ ID NO: 238)	27	15
6F	5'-CCAGGCCCTCTGATTCCTCACTGATTGCTC-3'	(SEQ ID NO: 239)	29	4
6R	5'-GCCACTGACAACCACCTTAACCCCTC-3'	(SEQ ID NO: 240)	27	29
7F	5'-GCCTCATCTTGGGCCTGTGTATCTCC-3'	(SEQ ID NO: 241)	27	3
7R	5'-GGCCAGTGTGCAAGGTGGCAAGTGGCTC-3'	(SEQ ID NO: 242)	28	5
8F	5'-GTAGGACCTGATTTCCTTACTGCCTCTTGC-3'	(SEQ ID NO: 243)	30	23
8R	5'-ATAACTGCACCCTTGGTCTCTCCACCGC-3'	(SEQ ID NO: 244)	29	20
9F	5'-CACTTTTATCACCTTTCCTTGCTCTTTCC-3'	(SEQ ID NO: 245)	30	3
9R	5'-AACTTTCCACTTGATAAGAGTCCCAAGAC-3'	(SEQ ID NO: 246)	30	7
10F	5'-ACTTACTTCTCCCTCTCTCTGTTGCTGC-3'	(SEQ ID NO: 247)	29	2
10R	5'-ATGGAATCCTATGGCTTTCCAACCTAGGAAG-3'	(SEQ ID NO: 248)	31	39
11F	5'-CATCTCTCCTCCCTGCTTCTGTCTCCTAC-3'	(SEQ ID NO: 249)	29	2
11R	5'-CTGACGCACACCTATTGCAAGCAAGGGTTC-3'	(SEQ ID NO: 250)	30	80

The term "INTRON" refers to the location in the intron where the primer anneals.

Alternatively the primers for exons 2 and 3 may be amplified together with primers:

p53-2/3F 5'GAAGCGTCTCATGCTGGAT-3' (SEQ ID NO: 251)
p53-2/3R 5'GGGGACTGTAGATGGGTGAA-3' (SEQ ID NO: 252)

For each exon analyzed, the following control PCRs are set up:

- (1) "Negative" DNA control (100 ng placental DNA (SIGMA CHEMICAL CO., St. Louis, Mo.)
- (2) Three "no template" controls

PCR for all exons is performed using the following thermocycling conditions:

Temperature	Time	Number of Cycles
95° C.	5 min (AMPLITAQ) or 10 min. (GOLD)	1
95° C.	30 sec.	} 30 cycles
55° C.	30 sec.	
72° C.	1 min	

36

-continued

Temperature	Time	Number of Cycles
72° C.	5 min.	1
4° C.	hold	1

Quality Control Agarose Gel of PCR Amplification:
The quality of the PCR products is examined prior to further analysis by electrophoresing an aliquot of each PCR

reaction sample on an agarose gel. 5 µl of each PCR reaction is run on an agarose gel along side a DNA 100 BP DNA

LADDER (Gibco BRL cat#15628-019). The electrophoresed PCR products are analyzed according to the following criteria:

Each patient sample must show a single band of the size corresponding the number of base pairs expected from the length of the PCR product from the forward primer to the reverse primer. If a patient sample demonstrates smearing or multiple bands, the PCR reaction must be repeated until a clean, single band is detected. If no PCR product is visible or if only a weak band is visible, but the control reactions with placental DNA template produced a robust band, the patient sample should be re-amplified with 2× as much template DNA.

All three "no template" reactions must show no amplification products. Any PCR product present in these reactions

US 6,951,721 B2

37

is the result of contamination. If any one of the "no template" reactions shows contamination, all PCR products should be discarded and the entire PCR set of reactions should be repeated after the appropriate PCR decontamination procedures have been taken.

The optimum amount of PCR product on the gel should be between 50 and 100 ng, which can be determined by comparing the intensity of the patient sample PCR products with that of the DNA ladder. If the patient sample PCR products contain less than 50 to 100 ng, the PCR reaction should be repeated until sufficient quantity is obtained.

DNA Sequencing

For DNA sequencing, double stranded PCR products are labeled with four different fluorescent dyes, one specific for each nucleotide, in a cycle sequencing reaction. With Dye Terminator Chemistry, when one of these nucleotides is incorporated into the elongating sequence it causes a termination at that point. Over the course of the cycle sequencing reaction, the dye-labeled nucleotides are incorporated along the length of the PCR product generating many different length fragments.

The dye-labeled PCR products will separate according to size when electrophoresed through a polyacrylamide gel. At the lower portion of the gel on an ABI automated sequencer, the fragments pass through a region where a laser beam continuously scans across the gel. The laser excites the fluorescent dyes attached to the fragments causing the emission of light at a specific wavelength for each dye. Either a photomultiplier tube (PMT) detects the fluorescent light and converts it into an electrical signal (ABI 373) or the light is collected and separated according to wavelength by a spectrograph onto a cooled, charge coupled device (CCD) camera (ABI 377). In either case the data collection software will collect the signals and store them for subsequent sequence analysis.

PCR products are first purified for sequencing using a QIAQUICK-SPIN PCR PURIFICATION KIT (QIAGEN #28104). The purified PCR products are labeled by adding primers, fluorescently tagged dNTPs and Taq Polymerase FS in an ABI Prism Dye Terminator Cycle Sequencing Kit (PERKIN ELMER/ABI catalog #02154) in a PERKIN ELMER GENEAMP 9600 thermocycler.

The amounts of each component are:

For Samples		For Controls	
Reagent	Volume	Reagent	Volume
Dye mix	8.0 μ L	PGEM	2.0 μ L
Primer (1.6 mM)	2.0 μ L	M13	2.0 μ L
PCR product	2.0 μ L	Dye mix	8.0 μ L
sdH ₂ O	8.0 μ L	sdH ₂ O	8.0 μ L

The thermocycling conditions are:

Temperature	Time	# of Cycles
96° C.	15 sec.	} 25
50° C.	5 sec.	
60° C.	4 min.	
4° C.	hold	1

The product is then loaded into a gel and placed into an ABI DNA Sequencer (Models 373A & 377) and run. The sequence obtained is analyzed by comparison to the wild type (reference) sequence using SEQUENCE NAVIGATOR

38

software. When a sequence does not align, it indicates a possible mutation or polymorphism. The DNA sequence is determined in both the forward and reverse directions. All results are provided to a second reader for review.

PCR Amplification for ASO

The genomic DNA is used as a template to amplify a separate DNA fragment encompassing the site of the mutation to be tested. The 50 μ L PCR reaction contains the following components: 1 μ L template (100 ng/ μ L) DNA, 5.0 μ L 10 \times PCR Buffer (PERKIN-ELMER), 2.5 μ L dNTP (2 mM each dATP, dCTP, dGTP, dTTP), 2.5 μ L Forward Primer (10 mM), 2.5 μ L Reverse Primer (10 μ M), 0.5 μ L (2.5 U total) AMPLITAQ® TAQ DNA POLYMERASE or AMPLITAQ GOLD™ DNA POLYMERASE (PERKIN-ELMER), 1.0 to 5.0 μ L (25 mM) MgCl₂ (depending on the primer) and distilled water (dH₂O) up to 50 μ L. All reagents for each exon except the genomic DNA can be combined in a master mix and aliquoted into the reaction tubes as a pooled mixture. The primers described above are used.

For each exon analyzed, the following control PCRs are set up:

- (1) "Negative" DNA control (100 ng placental DNA (SIGMA CHEMICAL CO., St. Louis, Mo.)
- (2) Three "no template" controls.

PCR for all exons is performed using the following thermocycling conditions:

Temperature	Time	Number of Cycles
95° C.	5 min (AMPLITAQ) or 10 min. (GOLD)	1
95° C.	30 sec.	} 30 cycles
55° C.	30 sec.	
72° C.	1 min	
72° C.	5 min.	1
4° C.	hold	1

The quality control agarose gel of PCR amplification is performed as above.

Binding PCR Products to Nylon Membrane

The PCR products are denatured no more than 30 minutes prior to binding the PCR products to the nylon membrane. To denature the PCR products, the remaining PCR reaction (45 μ L) and the appropriate positive control polymorphism gene amplification product are diluted to 200 μ L final volume with PCR Diluent Solution (500 mM NaOH, 2.0 M NaCl, 25 mM EDTA) and mixed thoroughly. The mixture is heated to 95° C. for 5 minutes, and immediately placed on ice and held on ice until loaded onto dot blotter, as described below.

The PCR products are bound to 9 cm by 13 cm nylon ZETA PROBE BLOTting MEMBRANE (BIO-RAD, Hercules, Calif., catalog number 162-0153) using a BIO-RAD dot blotter apparatus.

Pieces of 3MM filter paper [WHATMAN®, Clifton, N.J.] and nylon membrane are pre-wet in 10 \times SSC prepared fresh from 20 \times SSC buffer stock. The vacuum apparatus is rinsed thoroughly with dH₂O prior to assembly with the membrane. 100 μ L of each denatured PCR product is added to the wells of the blotting apparatus. Each row of the blotting apparatus contains a set of reactions for a single exon to be tested, including a placental DNA (negative) control, a synthetic oligonucleotide with the desired mutation or a PCR product from a known polymorphic sample (positive control), and three no template DNA controls.

After applying PCR products, the nylon filter is placed DNA side up on a piece of 3MM filter paper saturated with denaturing solution (1.5 M NaCl, 0.5 M NaOH) for 5

A000566

US 6,951,721 B2

39

minutes. The membrane is transferred to a piece of 3MM filter paper saturated with neutralizing solution (1 M Tris-HCl, pH 8, 1.5 M NaCl) for 5 minutes. The neutralized membrane is then transferred to a dry 3MM filter DNA side up, and exposed to ultraviolet light (STRALINKER, STRATAGENE, La Jolla, Calif.) for exactly 45 seconds to fix the DNA to the membrane. This UV crosslinking should be performed within 30 min. of the denaturation/neutralization steps. The nylon membrane is then cut into strips such that each strip contains a single row of blots of one set of reactions for a single exon.

Hybridizing Labeled Oligonucleotides to the Nylon Membrane Prehybridization

The strip is prehybridized at 52° C. incubation using the HYBAID® (SAVANT INSTRUMENTS, INC., Holbrook, N.Y.) hybridization oven. 2×SSC (15 to 20 ml) is preheated to 52° C. in a water bath. For each nylon strip, a single piece of nylon mesh cut slightly larger than the nylon membrane strip (approximately 1"×5") is pre-wet with 2×SSC. Each single nylon membrane is removed from the prehybridization solution and placed on top of the nylon mesh. The membrane/mesh "sandwich" is then transferred onto a piece of Parafilm™. The membrane/mesh sandwich is rolled lengthwise and placed into an appropriate HYBAID® bottle, such that the rotary action of the HYBAID® apparatus caused the membrane to unroll. The bottle is capped and gently rolled to cause the membrane/mesh to unroll and to evenly distribute the 2×SSC, making sure that no air bubbles formed between the membrane and mesh or between the mesh and the side of the bottle. The 2×SSC is discarded and replaced with 5 ml TMAC Hybridization Solution, which contains 3 M TMAC (tetramethyl ammoniumchloride-SIGMA T-3411), 100 mM Na₃PO₄ (pH 6.8), 1 mM EDTA, 5× Denhardt's (1% Ficoll, 1% polyvinylpyrrolidone, 1% BSA (fraction V)), 0.6% SDS, and 100 mg/ml Herring Sperm DNA. The filter strips are prehybridized at 52° C. with medium rotation (approx. 8.5 setting on the HYBAID® speed control) for at least one hour. Prehybridization can also be performed overnight.

Labeling Oligonucleotides

The DNA sequences of the numerous oligonucleotide probes are used to detect the p53 mutation. For each mutation, a polymorphic and a normal oligonucleotide must be labeled. While only five pairs of oligonucleotide probes are listed below, corresponding oligonucleotides for each mutation may be prepared and used in the same manner.

Polymorphism in Codon 21

wildtype 5'TTTTCAGACCTATGGAAAC-3' (SEQ ID NO: 253)
other wt 5'TTTCAGATCTATGGAAAC-3' (SEQ ID NO: 254)

Polymorphism in Codon 36

wildtype 5'CCCTTGCCGTCCTCAAGCA-3' (SEQ ID NO: 255)
other wt 5'CCCTTGCCATCTCAAGCA-3' (SEQ ID NO: 256)

Polymorphism in Codon 47

wildtype 5'CTGTCCCCGGACGATATT-3' (SEQ ID NO: 257)
other wt 5'CTGTCCCCAGACGATATT-3' (SEQ ID NO: 258)

40

Polymorphism in Codon 72

Polymorphism in Codon 213

wildtype 5'ACTTTTCGACATAGTGTG-3' (SEQ ID NO: 261)
other wt 5'ACTTTTCGGCATAGTGTG-3' (SEQ ID NO: 262)

Each labeling reaction contains 2 μl 5× Kinase buffer (or 1 μl of 10× Kinase buffer), 5 μl gamma-AIP ³²P (not more than one week old), 1 μl T4 polynucleotide kinase, 3 μl oligonucleotide (20 μM stock), sterile H₂O to 10 μl final volume if necessary. The reactions are incubated at 37° C. for 30 minutes, then at 65° C. for 10 minutes to heat inactivate the kinase. The kinase reaction is diluted with an equal volume (10 μl) of sterile dH₂O (distilled water).

The oligonucleotides are purified on STE MICRO SELECT-D, G-25 spin columns (catalog no. 5303-356769), according to the manufacturer's instructions. The 20 μl synthetic oligonucleotide eluate is diluted with 80 μl dH₂O (final volume=100 μl). The amount of radioactivity in the oligonucleotide sample is determined by measuring the radioactive counts per minute (cpm). The total radioactivity must be at least 2 million cpm. For any samples containing less than 2 million cpm total, the labeling reaction is repeated.

Hybridization with Oligonucleotides

Approximately 2–5 million cpm of the labeled polymorphic oligonucleotide probe is diluted into 5 ml of TMAC hybridization solution, containing 40 μl of 20 μM stock of unlabeled normal oligonucleotide. The probe mix is preheated to 52° C. in the hybridization oven. The prehybridization solution is removed from each bottle and replaced with the probe mix. The filter is hybridized for 1 hour at 52° C. with moderate agitation. Following hybridization, the probe mix is decanted into a storage tube and stored at –20° C. The filter is rinsed by adding approximately 20 ml of 2×SSC+0.1% SDS at room temperature and rolling the capped bottle gently for approximately 30 seconds and pouring off the rinse. The filter is then washed with 2×SSC+0.1% SDS at room temperature for 20 to 30 minutes, with shaking.

The membrane is removed from the wash and placed on a dry piece of 3MM WHATMAN filter paper then wrapped in one layer of plastic wrap, placed on the autoradiography film, and exposed for about five hours depending upon a survey meter indicating the level of radioactivity. The film is developed in an automatic film processor.

Control Hybridization with Normal Oligonucleotides

The purpose of this step is to ensure that the PCR products are transferred efficiently to the nylon membrane.

Following hybridization with the polymorphic oligonucleotide each nylon membrane is washed in 2×SSC, 0.1% SDS for 20 minutes at 65° C. to melt off the polymorphic oligonucleotide probes. The nylon strips are then prehybridized together in 40 ml of TMAC hybridization solution for at least 1 hour at 52° C. in a shaking water bath. 2–5 million counts of each of the normal labeled oligonucleotide probes plus 40 ml of 20 mM stock of unlabeled normal oligonucleotide are added directly to the container containing the nylon membranes and the prehybridization solution. The filter and probes are hybridized at 52° C. with shaking for at least 1 hour. Hybridization can be performed overnight, if necessary. The hybridization solution is poured off, and the nylon membrane is rinsed in 2×SSC, 0.1% SDS for 1 minute with gentle swirling by hand. The rinse is poured off and the membrane is washed in 2×SSC, 0.1% SDS at room temperature for 20 minutes with shaking.

A000567

US 6,951,721 B2

41

The nylon membrane is removed placed on a dry piece of 3MM WHATMAN filter paper. The nylon membrane is then wrapped in one layer of plastic wrap and placed on autoradiography film, and exposure is for at least 1 hour.

For each sample, adequate transfer to the membrane is indicated by a strong autoradiographic hybridization signal. For each sample, an absent or weak signal when hybridized with its normal oligonucleotide, indicates an unsuccessful transfer of PCR product, and it is a false negative. The ASO analysis must be repeated for any sample that did not successfully transfer to the nylon membrane.

Homozygous individuals having haplotypes with the single nucleotide polymorphism (SNP) arginine at codon 72 are overrepresented in the genomic alleles of cervical cancer patients. In addition, it was recently published that cervical tumors have the SNP arginine at codon 72 at significantly higher frequency than normal tissue. Storey et al, *Nature*, 393: 229-234 (1998). Healthy women having such haplotypes are candidates for the HPV vaccines to prevent HPV infection, treat venereal warts, treat cervical and other related cancers, and prevent reoccurrence of venereal warts previously treated.

EXAMPLE 7

Pharmacogenetic Analysis of P1Haplotype and Platelet Sensitivity to Aspirin

Aspirin has been a standard anticoagulant therapy for patients who have had a heart attack. In recent years, aspirin therapy has been extended to individuals with a history or at risk for stroke (apoplexy) and phlebitis. It has even been proposed that every person over 50 years of age should take aspirin.

However, some people cannot take aspirin due to allergy, erosion of the stomach lining etc. Furthermore, research has shown that aspirin prevents heart attacks in about 40 percent of patients taking aspirin. Thus, it is desirable to determine which people will respond to aspirin and which will not in order to administer other anticoagulant or antiplatelet medication.

Platelet aggregation is recognized as an important step in the formation of a blockage which will cause a myocardial infarction and unstable angina. Platelet aggregation is based on glycoprotein gpIIb/IIIa. Different forms of this glycoprotein have been known. Weiss et al, *Tissue Antigens*, 46: 374-381 (1995), Kunicki et al, *Molecular Immunology* 16: 353-60 (1979). Methods for determining various polymor-

42

phisms may be done by DNA analysis. Newman et al, *Journal of Clinical Investigation* 83:1778-81 (1989). It has been reported that patients having one polymorphic form of the PI gene have a higher incidence for acute coronary thrombosis, particularly in patients younger than 60. Weiss et al, *New England Journal of Medicine* 334(17): 1090-1094 (1996). However, these findings were contradicted by Ridker, et al, *Lancet* 349: 385-388 (1997) with comments in *Lancet* on pages 370-371, 1099-1100 and 1100-1. Adding to the debate, it was recently published that platelet aggregation from haplotype PI^{A2} containing individuals are less inhibited by aspirin at certain concentrations than individuals homozygous for haplotype PI^{A1}. Cooke et al, *Lancet* 351: 1253 (1998).

Resolving the issue for people at risk of heart attacks, stroke and other thrombotic disorders is desirable, particularly in distinguishing between those who can take aspirin or who should take other medication which is more costly and with greater side effects.

Experimental Protocol

Blood samples are taken from 50 healthy individuals ages 50-55. Family history and personal histories of heart disease and other thrombotic disorders are recorded. White blood cells are collected and genomic DNA is extracted from the white blood cells, PCR amplified and the sequence determined by ASO or sequenced as in the Examples above using different primers and probes. Newman et al., *Journal of Clinical Investigation* 83:1778-81 (1989). As before, PCR primers and ASO probes are designed to type these individuals for exon 2 to determine which base exists at nucleotide position 1565: a T or a C. at the amino acid level, codon 33 is changed from a leucine to a proline.

Individuals having haplotype PI^{A2} either in homozygous or heterozygous form are instructed to either take high dosages of aspirin (2000 mg per day) or not take aspirin and given other medication appropriate for their individual needs. Individuals homozygous for haplotype PI^{A1} are instructed to take aspirin at low dosages (350 mg per day).

The present invention is not to be limited in scope by the specific embodiments described herein. Indeed, various modifications of the invention in addition to those described herein will become apparent to those skilled in the art from the foregoing description and accompanying figure. Such modifications are intended to fall within the scope of the appended claims.

Various publications are cited herein, the disclosures of which are incorporated by reference in their entireties.

SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 268

<210> SEQ ID NO 1

<211> LENGTH: 19

<212> TYPE: DNA

<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 1

cgcgctctgct tatgattgg

19

<210> SEQ ID NO 2

<211> LENGTH: 18

<212> TYPE: DNA

A000568

US 6,951,721 B2

45

46

-continued

<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 10

tttttttttt ttttacctga 20

<210> SEQ ID NO 11
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 11

actaatgagc ttgccattct 20

<210> SEQ ID NO 12
<211> LENGTH: 19
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 12

tgggtaactg caggttaca 19

<210> SEQ ID NO 13
<211> LENGTH: 18
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 13

gacttacgtg cttagtgtg 18

<210> SEQ ID NO 14
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 14

agtatatatt gtatgagttg aagg 24

<210> SEQ ID NO 15
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 15

gatttggtatt ctgtaaaatg agatc 25

<210> SEQ ID NO 16
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 16

ggcctttgct ttttaaaaat aac 23

<210> SEQ ID NO 17
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 17

gtctttaccc attatttata gg 22

<210> SEQ ID NO 18

A000570

US 6,951,721 B2

47

48

-continued

<211> LENGTH: 22
 <212> TYPE: DNA
 <213> ORGANISM: Homo sapiens

<400> SEQUENCE: 18

gtatagacaa aagaattatt cc 22

<210> SEQ ID NO 19
 <211> LENGTH: 23
 <212> TYPE: DNA
 <213> ORGANISM: Homo sapiens

<400> SEQUENCE: 19

ggtagtaggt atttatggaa tac 23

<210> SEQ ID NO 20
 <211> LENGTH: 19
 <212> TYPE: DNA
 <213> ORGANISM: Homo sapiens

<400> SEQUENCE: 20

catgttagag catttaggg 19

<210> SEQ ID NO 21
 <211> LENGTH: 19
 <212> TYPE: DNA
 <213> ORGANISM: Homo sapiens

<400> SEQUENCE: 21

cacattgctt ctagtacac 19

<210> SEQ ID NO 22
 <211> LENGTH: 18
 <212> TYPE: DNA
 <213> ORGANISM: Homo sapiens

<400> SEQUENCE: 22

ccaggtgaca ttcagaac 18

<210> SEQ ID NO 23
 <211> LENGTH: 19
 <212> TYPE: DNA
 <213> ORGANISM: Homo sapiens

<400> SEQUENCE: 23

attcagtatt cctgtgtac 19

<210> SEQ ID NO 24
 <211> LENGTH: 17
 <212> TYPE: DNA
 <213> ORGANISM: Homo sapiens

<400> SEQUENCE: 24

cgttaccccc acaaagc 17

<210> SEQ ID NO 25
 <211> LENGTH: 20
 <212> TYPE: DNA
 <213> ORGANISM: Homo sapiens

<400> SEQUENCE: 25

atgctatgtc agtgtaaacc 20

A000571

US 6,951,721 B2

49

50

-continued

```

<210> SEQ ID NO 26
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 26

ccacaggaaa acaactatta                20

<210> SEQ ID NO 27
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 27

taccacattt tatgtgatgg                20

<210> SEQ ID NO 28
<211> LENGTH: 17
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 28

ggggtagtaa gtttccc                  17

<210> SEQ ID NO 29
<211> LENGTH: 18
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 29

cttttctcat gctgtccc                  18

<210> SEQ ID NO 30
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 30

atagagaagc taagttaaac                20

<210> SEQ ID NO 31
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 31

taattactca tgggacattc                20

<210> SEQ ID NO 32
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 32

ggcactgaca gttaacacta                20

<210> SEQ ID NO 33
<211> LENGTH: 18
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: M13 tail for forward primers

<400> SEQUENCE: 33

```

A000572

US 6,951,721 B2

51

52

-continued

 tgtaaaacga cggccagt 18

<210> SEQ ID NO 34
 <211> LENGTH: 18
 <212> TYPE: DNA
 <213> ORGANISM: Artificial sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: M13 tail for reverse primers

<400> SEQUENCE: 34
 caggaaacag ctatgacc 18

<210> SEQ ID NO 35
 <211> LENGTH: 19
 <212> TYPE: DNA
 <213> ORGANISM: Homo sapiens
 <400> SEQUENCE: 35
 aggcaactgag gtgattggc 19

<210> SEQ ID NO 36
 <211> LENGTH: 19
 <212> TYPE: DNA
 <213> ORGANISM: Homo sapiens
 <400> SEQUENCE: 36
 tcgtagccct taagtgagc 19

<210> SEQ ID NO 37
 <211> LENGTH: 22
 <212> TYPE: DNA
 <213> ORGANISM: Homo sapiens
 <400> SEQUENCE: 37
 tgaggcacta ttgtttgtat tt 22

<210> SEQ ID NO 38
 <211> LENGTH: 22
 <212> TYPE: DNA
 <213> ORGANISM: Homo sapiens
 <400> SEQUENCE: 38
 tgttggtgtt gaatttttca gt 22

<210> SEQ ID NO 39
 <211> LENGTH: 22
 <212> TYPE: DNA
 <213> ORGANISM: Homo sapiens
 <400> SEQUENCE: 39
 agagatttgg aaaatgagta ac 22

<210> SEQ ID NO 40
 <211> LENGTH: 19
 <212> TYPE: DNA
 <213> ORGANISM: Homo sapiens
 <400> SEQUENCE: 40
 acaatgtcat cacaggagg 19

<210> SEQ ID NO 41
 <211> LENGTH: 18
 <212> TYPE: DNA
 <213> ORGANISM: Homo sapiens

A000573

US 6,951,721 B2

53

54

-continued

```

<400> SEQUENCE: 41
tgaggtgaca gtgggtga                                18

<210> SEQ ID NO 42
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 42
gattactctg agacctaggc                                20

<210> SEQ ID NO 43
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 43
gattttctct tttcccttg gg                              22

<210> SEQ ID NO 44
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 44
caaacaaagc ttcaacaatt tac                            23

<210> SEQ ID NO 45
<211> LENGTH: 26
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 45
gggttttatt ttcaagtact tctatg                        26

<210> SEQ ID NO 46
<211> LENGTH: 26
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 46
gctcagcaac tgttcaatgt atgagc                        26

<210> SEQ ID NO 47
<211> LENGTH: 18
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 47
ctagtgtgtg tttttggc                                18

<210> SEQ ID NO 48
<211> LENGTH: 18
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 48
cataacctta tctccacc                                18

<210> SEQ ID NO 49
<211> LENGTH: 23
<212> TYPE: DNA

```

A000574

US 6,951,721 B2

57

58

-continued

<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 57

ttaaatacag actttgctac 20

<210> SEQ ID NO 58
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 58

gaaaagccaa agttagaagg 20

<210> SEQ ID NO 59
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 59

tgcaaccac aaaatttggc 20

<210> SEQ ID NO 60
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 60

ctttctccat ttccaaaacc 20

<210> SEQ ID NO 61
<211> LENGTH: 18
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 61

tggtgtctct agttctgg 18

<210> SEQ ID NO 62
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 62

cattgttgta gtagctctgc 20

<210> SEQ ID NO 63
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 63

gcagaactat gtctgtctca t 21

<210> SEQ ID NO 64
<211> LENGTH: 19
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 64

cggtcagttg aaatgtcag 19

<210> SEQ ID NO 65

A000576

US 6,951,721 B2

59

60

-continued

<211> LENGTH: 19
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 65
catttggatc cgtaaagc 19

<210> SEQ ID NO 66
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 66
caccggctg gaaatttat ttg 23

<210> SEQ ID NO 67
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 67
ggaaaggcac tggagaaatg gg 22

<210> SEQ ID NO 68
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 68
ccctccagca cacatgcatg taccg 25

<210> SEQ ID NO 69
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 69
taagtagtct gtgatctccg 20

<210> SEQ ID NO 70
<211> LENGTH: 18
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 70
atgtatgagg tcctgtcc 18

<210> SEQ ID NO 71
<211> LENGTH: 18
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 71
gacaccagtg tatgttgg 18

<210> SEQ ID NO 72
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 72
gagaaagaag aacacatccc 20

A000577

US 6,951,721 B2

61

62

-continued

<210> SEQ ID NO 73
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 73

gaagttgtca tttataaac cttt 24

<210> SEQ ID NO 74
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 74

tgtctttttct tccctagtat gt 22

<210> SEQ ID NO 75
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 75

tcctgacaca gcagacatta 20

<210> SEQ ID NO 76
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 76

ttggatttcg ttctcactta 20

<210> SEQ ID NO 77
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 77

ctcttaaggc cagttgtgag 20

<210> SEQ ID NO 78
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 78

ttcctactgt ggttgcttcc 20

<210> SEQ ID NO 79
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 79

cttatttttag tgccttaaa agg 23

<210> SEQ ID NO 80
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 80

tttcatggac agcacttgag tg 22

A000578

US 6,951,721 B2

63

64

-continued

<210> SEQ ID NO 81
 <211> LENGTH: 23
 <212> TYPE: DNA
 <213> ORGANISM: Homo sapiens

<400> SEQUENCE: 81

cacaacaaag agcatacata ggg

23

<210> SEQ ID NO 82
 <211> LENGTH: 20
 <212> TYPE: DNA
 <213> ORGANISM: Homo sapiens

<400> SEQUENCE: 82

tcgggttcac tctgtagaag

20

<210> SEQ ID NO 83
 <211> LENGTH: 21
 <212> TYPE: DNA
 <213> ORGANISM: Homo sapiens

<400> SEQUENCE: 83

ttctcttcag gaggaaaagc a

21

<210> SEQ ID NO 84
 <211> LENGTH: 21
 <212> TYPE: DNA
 <213> ORGANISM: Homo sapiens

<400> SEQUENCE: 84

gctgcctacc acaaatacaa a

21

<210> SEQ ID NO 85
 <211> LENGTH: 21
 <212> TYPE: DNA
 <213> ORGANISM: Homo sapiens

<400> SEQUENCE: 85

ccacagtaga tgctcagtaa a

21

<210> SEQ ID NO 86
 <211> LENGTH: 23
 <212> TYPE: DNA
 <213> ORGANISM: Homo sapiens

<400> SEQUENCE: 86

taggaaaata ccagcttcac aga

23

<210> SEQ ID NO 87
 <211> LENGTH: 20
 <212> TYPE: DNA
 <213> ORGANISM: Homo sapiens

<400> SEQUENCE: 87

tggtcagctt tctgtaatcg

20

<210> SEQ ID NO 88
 <211> LENGTH: 24
 <212> TYPE: DNA
 <213> ORGANISM: Homo sapiens

<400> SEQUENCE: 88

gtatctaccc actctcttct tcag

24

A000579

US 6,951,721 B2

65

66

-continued

<210> SEQ ID NO 89
<211> LENGTH: 19
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 89

ccacctccaa ggtgtatca

19

<210> SEQ ID NO 90
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 90

tgttatgttg gctccttgct

20

<210> SEQ ID NO 91
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 91

cactaaagac agaatgaatc ta

22

<210> SEQ ID NO 92
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 92

gaagaaccag aatattcatc ta

22

<210> SEQ ID NO 93
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 93

tgatggggag tctgaatcaa

20

<210> SEQ ID NO 94
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 94

tctgctttct tgataaaatc ct

22

<210> SEQ ID NO 95
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 95

agcgtcccct cacaaataaa

20

<210> SEQ ID NO 96
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 96

A000580

US 6,951,721 B2

67

68

-continued

tcaagcgcat gaatatgcct	20
<210> SEQ ID NO 97 <211> LENGTH: 22 <212> TYPE: DNA <213> ORGANISM: Homo sapiens <400> SEQUENCE: 97	
gtataagcaa tatggaactc ga	22
<210> SEQ ID NO 98 <211> LENGTH: 23 <212> TYPE: DNA <213> ORGANISM: Homo sapiens <400> SEQUENCE: 98	
ttaagttcac tggatttga aca	23
<210> SEQ ID NO 99 <211> LENGTH: 20 <212> TYPE: DNA <213> ORGANISM: Homo sapiens <400> SEQUENCE: 99	
gacagcgata ctttcccaga	20
<210> SEQ ID NO 100 <211> LENGTH: 21 <212> TYPE: DNA <213> ORGANISM: Homo sapiens <400> SEQUENCE: 100	
tggaacaacc atgaattagt c	21
<210> SEQ ID NO 101 <211> LENGTH: 20 <212> TYPE: DNA <213> ORGANISM: Homo sapiens <400> SEQUENCE: 101	
ggaagttagc actctaggga	20
<210> SEQ ID NO 102 <211> LENGTH: 22 <212> TYPE: DNA <213> ORGANISM: Homo sapiens <400> SEQUENCE: 102	
gcagtgatat taactgtctg ta	22
<210> SEQ ID NO 103 <211> LENGTH: 22 <212> TYPE: DNA <213> ORGANISM: Homo sapiens <400> SEQUENCE: 103	
tgggtcctta aagaacaaaa gt	22
<210> SEQ ID NO 104 <211> LENGTH: 21 <212> TYPE: DNA <213> ORGANISM: Homo sapiens <400> SEQUENCE: 104	

A000581

US 6,951,721 B2

69

70

-continued

tcaggtgaca ttgaatcttc c	21
<210> SEQ ID NO 105 <211> LENGTH: 21 <212> TYPE: DNA <213> ORGANISM: Homo sapiens <400> SEQUENCE: 105	
ccactttttc ccatcaagtc a	21
<210> SEQ ID NO 106 <211> LENGTH: 22 <212> TYPE: DNA <213> ORGANISM: Homo sapiens <400> SEQUENCE: 106	
tcaggatgct tacaattact tc	22
<210> SEQ ID NO 107 <211> LENGTH: 23 <212> TYPE: DNA <213> ORGANISM: Homo sapiens <400> SEQUENCE: 107	
caaaattgaa tgctatgctt aga	23
<210> SEQ ID NO 108 <211> LENGTH: 20 <212> TYPE: DNA <213> ORGANISM: Homo sapiens <400> SEQUENCE: 108	
tcggtaaccc tgagccaaat	20
<210> SEQ ID NO 109 <211> LENGTH: 20 <212> TYPE: DNA <213> ORGANISM: Homo sapiens <400> SEQUENCE: 109	
gcaaaagcgt ccagaaagga	20
<210> SEQ ID NO 110 <211> LENGTH: 22 <212> TYPE: DNA <213> ORGANISM: Homo sapiens <400> SEQUENCE: 110	
tatttgcagt caagtcttcc aa	22
<210> SEQ ID NO 111 <211> LENGTH: 21 <212> TYPE: DNA <213> ORGANISM: Homo sapiens <400> SEQUENCE: 111	
gtaatattgg caaaggcatc t	21
<210> SEQ ID NO 112 <211> LENGTH: 22 <212> TYPE: DNA <213> ORGANISM: Homo sapiens	

A000582

US 6,951,721 B2

71

72

-continued

<400> SEQUENCE: 112

taaaatgtgc tccccaaaag ca	22
--------------------------	----

<210> SEQ ID NO 113

<211> LENGTH: 20

<212> TYPE: DNA

<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 113

gtcctgccaa tgagaagaaa	20
-----------------------	----

<210> SEQ ID NO 114

<211> LENGTH: 21

<212> TYPE: DNA

<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 114

tgtcagcaaa cctaagaatg t	21
-------------------------	----

<210> SEQ ID NO 115

<211> LENGTH: 21

<212> TYPE: DNA

<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 115

aatggaaagc ttctcaaagt a	21
-------------------------	----

<210> SEQ ID NO 116

<211> LENGTH: 21

<212> TYPE: DNA

<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 116

atgttgagc taggtcctta c	21
------------------------	----

<210> SEQ ID NO 117

<211> LENGTH: 22

<212> TYPE: DNA

<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 117

ctaacctgaa ttatcactat ca	22
--------------------------	----

<210> SEQ ID NO 118

<211> LENGTH: 21

<212> TYPE: DNA

<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 118

gtgtataaat gcctgtatgc a	21
-------------------------	----

<210> SEQ ID NO 119

<211> LENGTH: 19

<212> TYPE: DNA

<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 119

tggctgcccc ggaagtatg	19
----------------------	----

<210> SEQ ID NO 120

<211> LENGTH: 23

<212> TYPE: DNA

<213> ORGANISM: Homo sapiens

A000583

US 6,951,721 B2

73

74

-continued

```

<400> SEQUENCE: 120
aaccagaata tctttatgta gga                23

<210> SEQ ID NO 121
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 121
aattcttaac agagaccaga ac                22

<210> SEQ ID NO 122
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 122
aaaactcttt ccagaatggt gt                22

<210> SEQ ID NO 123
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 123
gtgtagaacg tgcaggattg                20

<210> SEQ ID NO 124
<211> LENGTH: 18
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 124
tcgcctcatg tggtttta                18

<210> SEQ ID NO 125
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 125
ggctcttttag cttcttagga c                21

<210> SEQ ID NO 126
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 126
gagaccattt tccagcatc                20

<210> SEQ ID NO 127
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 127
ctgtcattct tcctgtgctc                20

<210> SEQ ID NO 128
<211> LENGTH: 21
<212> TYPE: DNA

```

A000584

US 6,951,721 B2

77

78

-continued

```
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
```

<400> SEQUENCE: 136

catttagcc attcattcaa caa

23

```
<210> SEQ ID NO 137
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
```

<400> SEQUENCE: 137

atgaattgac actaatctct gc

22

```
<210> SEQ ID NO 138
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
```

<400> SEQUENCE: 138

gtagccagga cagtagaagg a

21

```
<210> SEQ ID NO 139
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
```

<400> SEQUENCE: 139

gaataatata aattatatgg cttta

24

```
<210> SEQ ID NO 140
<211> LENGTH: 19
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
```

<400> SEQUENCE: 140

cctagtcttg ctagttctt

19

```
<210> SEQ ID NO 141
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
```

<400> SEQUENCE: 141

atctgaagtg gaaccaaattg atac

24

```
<210> SEQ ID NO 142
<211> LENGTH: 26
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
```

<400> SEQUENCE: 142

acgtggcaaa gaattctctg aagtaa

26

```
<210> SEQ ID NO 143
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
```

<400> SEQUENCE: 143

aagaagcaaa atgtaataag ga

22

<210> SEQ ID NO 144

A000586

US 6,951,721 B2

79

80

-continued

```
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 144
catttaaagc acatacatct tg                22

<210> SEQ ID NO 145
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 145
tctagaggca aagaatcata c                21

<210> SEQ ID NO 146
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 146
caagattatt cctttcatta gc                22

<210> SEQ ID NO 147
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 147
aaccaaaaca caaatctaag ag                22

<210> SEQ ID NO 148
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 148
gtcattttta tatgctgctt tac                23

<210> SEQ ID NO 149
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 149
ggttttatat ggagacacag g                21

<210> SEQ ID NO 150
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 150
gtatttacaa tttcaacaca agc                23

<210> SEQ ID NO 151
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 151
atcacagttt tggaggtagc                20
```

A000587

US 6,951,721 B2

81

82

-continued

```

<210> SEQ ID NO 152
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 152

ctgacttcct gattcttcta a                21

<210> SEQ ID NO 153
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 153

accatgtagc aaatgagggt ct              22

<210> SEQ ID NO 154
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 154

gcttttgtct gttttcctcc aa              22

<210> SEQ ID NO 155
<211> LENGTH: 19
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 155

aaccacaccc ttaagatga                  19

<210> SEQ ID NO 156
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 156

gcataagtag tggattttgc                  20

<210> SEQ ID NO 157
<211> LENGTH: 31
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 157

gggcccggat ccatggcaat gcagatgcag c    31

<210> SEQ ID NO 158
<211> LENGTH: 43
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 158

gggcccacat ggatatcatt cagtctttgg catctcccac tcc 43

<210> SEQ ID NO 159
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 159

cacgaggcat ggcgctgagg                  20

```

A000588

US 6,951,721 B2

83

84

-continued

<210> SEQ ID NO 160
 <211> LENGTH: 20
 <212> TYPE: DNA
 <213> ORGANISM: Homo sapiens

<400> SEQUENCE: 160

ccgggccttg tctgtccact 20

<210> SEQ ID NO 161
 <211> LENGTH: 20
 <212> TYPE: DNA
 <213> ORGANISM: Homo sapiens

<400> SEQUENCE: 161

gtctacccca ttgaccatgg 20

<210> SEQ ID NO 162
 <211> LENGTH: 20
 <212> TYPE: DNA
 <213> ORGANISM: Homo sapiens

<400> SEQUENCE: 162

tcacatctcg agtactgctg 20

<210> SEQ ID NO 163
 <211> LENGTH: 20
 <212> TYPE: DNA
 <213> ORGANISM: Homo sapiens

<400> SEQUENCE: 163

tgcaggagga agaagacctg 20

<210> SEQ ID NO 164
 <211> LENGTH: 20
 <212> TYPE: DNA
 <213> ORGANISM: Homo sapiens

<400> SEQUENCE: 164

tctgtcagcg ccaggggact 20

<210> SEQ ID NO 165
 <211> LENGTH: 20
 <212> TYPE: DNA
 <213> ORGANISM: Homo sapiens

<400> SEQUENCE: 165

agcacaggcc tgctgcacct 20

<210> SEQ ID NO 166
 <211> LENGTH: 20
 <212> TYPE: DNA
 <213> ORGANISM: Homo sapiens

<400> SEQUENCE: 166

gaaaagggga agtggggcag 20

<210> SEQ ID NO 167
 <211> LENGTH: 32
 <212> TYPE: DNA
 <213> ORGANISM: Homo sapiens

<400> SEQUENCE: 167

agcccaggcc ccaacacagc cccatggcct ct 32

A000589

US 6,951,721 B2

85

86

-continued

<210> SEQ ID NO 168
<211> LENGTH: 32
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 168

cttaggagag ttttattcat tcattgatcc ag

32

<210> SEQ ID NO 169
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 169

aacagtacaa tgactgggct c

21

<210> SEQ ID NO 170
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 170

tcagcgcttc tgcacacagt

20

<210> SEQ ID NO 171
<211> LENGTH: 17
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 171

acatgacagc gatactt

17

<210> SEQ ID NO 172
<211> LENGTH: 17
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 172

acatgacagt gatactt

17

<210> SEQ ID NO 173
<211> LENGTH: 17
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 173

agtatttcat tgggtacc

17

<210> SEQ ID NO 174
<211> LENGTH: 17
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 174

agtatttcac tgggtacc

17

<210> SEQ ID NO 175
<211> LENGTH: 17
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 175

A000590

US 6,951,721 B2

87

88

-continued

catttgctcc gttttca	17
<210> SEQ ID NO 176	
<211> LENGTH: 17	
<212> TYPE: DNA	
<213> ORGANISM: Homo sapiens	
<400> SEQUENCE: 176	
catttgctct gttttca	17
<210> SEQ ID NO 177	
<211> LENGTH: 17	
<212> TYPE: DNA	
<213> ORGANISM: Homo sapiens	
<400> SEQUENCE: 177	
tttttaaaga agccagc	17
<210> SEQ ID NO 178	
<211> LENGTH: 17	
<212> TYPE: DNA	
<213> ORGANISM: Homo sapiens	
<400> SEQUENCE: 178	
tttttaaagg agccagc	17
<210> SEQ ID NO 179	
<211> LENGTH: 17	
<212> TYPE: DNA	
<213> ORGANISM: Homo sapiens	
<400> SEQUENCE: 179	
gcgtccagaa aggagag	17
<210> SEQ ID NO 180	
<211> LENGTH: 17	
<212> TYPE: DNA	
<213> ORGANISM: Homo sapiens	
<400> SEQUENCE: 180	
gcgtccagag aggagag	17
<210> SEQ ID NO 181	
<211> LENGTH: 17	
<212> TYPE: DNA	
<213> ORGANISM: Homo sapiens	
<400> SEQUENCE: 181	
aagtgactct tctgccc	17
<210> SEQ ID NO 182	
<211> LENGTH: 17	
<212> TYPE: DNA	
<213> ORGANISM: Homo sapiens	
<400> SEQUENCE: 182	
aagtgactcc tctgccc	17
<210> SEQ ID NO 183	
<211> LENGTH: 17	
<212> TYPE: DNA	
<213> ORGANISM: Homo sapiens	
<400> SEQUENCE: 183	

A000591

US 6,951,721 B2

89

90

-continued

tgtgcccaga gtccagc	17
<210> SEQ ID NO 184 <211> LENGTH: 17 <212> TYPE: DNA <213> ORGANISM: Homo sapiens <400> SEQUENCE: 184	
tgtgcccagg gtccagc	17
<210> SEQ ID NO 185 <211> LENGTH: 17 <212> TYPE: DNA <213> ORGANISM: Homo sapiens <400> SEQUENCE: 185	
ggaataagca gaaactg	17
<210> SEQ ID NO 186 <211> LENGTH: 17 <212> TYPE: DNA <213> ORGANISM: Homo sapiens <400> SEQUENCE: 186	
ggaataagcg gaaactg	17
<210> SEQ ID NO 187 <211> LENGTH: 17 <212> TYPE: DNA <213> ORGANISM: Homo sapiens <400> SEQUENCE: 187	
aaaagacatg acagcga	17
<210> SEQ ID NO 188 <211> LENGTH: 17 <212> TYPE: DNA <213> ORGANISM: Homo sapiens <400> SEQUENCE: 188	
aaaagacata acagcga	17
<210> SEQ ID NO 189 <211> LENGTH: 17 <212> TYPE: DNA <213> ORGANISM: Homo sapiens <400> SEQUENCE: 189	
aagaagccag ctcaagc	17
<210> SEQ ID NO 190 <211> LENGTH: 17 <212> TYPE: DNA <213> ORGANISM: Homo sapiens <400> SEQUENCE: 190	
aagaagccaa ctcaagc	17
<210> SEQ ID NO 191 <211> LENGTH: 17 <212> TYPE: DNA <213> ORGANISM: Homo sapiens	

A000592

US 6,951,721 B2

91

92

-continued

<400> SEQUENCE: 191

catgacagtg atacttt

17

<210> SEQ ID NO 192

<211> LENGTH: 17

<212> TYPE: DNA

<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 192

catgacagta atacttt

17

<210> SEQ ID NO 193

<211> LENGTH: 17

<212> TYPE: DNA

<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 193

taggacattg gcattga

17

<210> SEQ ID NO 194

<211> LENGTH: 18

<212> TYPE: DNA

<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 194

taggacatgt ggcattga

18

<210> SEQ ID NO 195

<211> LENGTH: 18

<212> TYPE: DNA

<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 195

cttctgattt gctacatt

18

<210> SEQ ID NO 196

<211> LENGTH: 18

<212> TYPE: DNA

<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 196

cttctgatgt gctacatt

18

<210> SEQ ID NO 197

<211> LENGTH: 18

<212> TYPE: DNA

<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 197

ggcttctctg attttggt

18

<210> SEQ ID NO 198

<211> LENGTH: 18

<212> TYPE: DNA

<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 198

ggcttctcgg attttggt

18

<210> SEQ ID NO 199

<211> LENGTH: 18

<212> TYPE: DNA

<213> ORGANISM: Homo sapiens

A000593

US 6,951,721 B2

93

94

-continued

<400> SEQUENCE: 199

ttttgaatat tgtactgg

18

<210> SEQ ID NO 200

<211> LENGTH: 18

<212> TYPE: DNA

<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 200

ttttgaatgt tgtactgg

18

<210> SEQ ID NO 201

<211> LENGTH: 18

<212> TYPE: DNA

<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 201

attagctact tggaagac

18

<210> SEQ ID NO 202

<211> LENGTH: 18

<212> TYPE: DNA

<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 202

attagctatt tggaagac

18

<210> SEQ ID NO 203

<211> LENGTH: 18

<212> TYPE: DNA

<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 203

ccatttggtc atgtaatc

18

<210> SEQ ID NO 204

<211> LENGTH: 18

<212> TYPE: DNA

<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 204

ccatttgccc atgtaatc

18

<210> SEQ ID NO 205

<211> LENGTH: 18

<212> TYPE: DNA

<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 205

tagcttggtt ttctaaac

18

<210> SEQ ID NO 206

<211> LENGTH: 18

<212> TYPE: DNA

<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 206

tagcttggtt ttctaaac

18

<210> SEQ ID NO 207

<211> LENGTH: 18

<212> TYPE: DNA

A000594

US 6,951,721 B2

95

96

-continued

<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 207

attgaacaa cagaatca 18

<210> SEQ ID NO 208

<211> LENGTH: 18

<212> TYPE: DNA

<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 208

attgaacga cagaatca 18

<210> SEQ ID NO 209

<211> LENGTH: 18

<212> TYPE: DNA

<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 209

tgaaaatgtg atttagtt 18

<210> SEQ ID NO 210

<211> LENGTH: 18

<212> TYPE: DNA

<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 210

tgaaaatgcg atttagtt 18

<210> SEQ ID NO 211

<211> LENGTH: 18

<212> TYPE: DNA

<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 211

ttccatggcc ttctaata 18

<210> SEQ ID NO 212

<211> LENGTH: 18

<212> TYPE: DNA

<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 212

ttccatggtc ttctaata 18

<210> SEQ ID NO 213

<211> LENGTH: 18

<212> TYPE: DNA

<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 213

cttgaaggcg tacaacag 18

<210> SEQ ID NO 214

<211> LENGTH: 18

<212> TYPE: DNA

<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 214

cttgaagggtg tacaacag 18

<210> SEQ ID NO 215

<211> LENGTH: 19

A000595

US 6,951,721 B2

97

98

-continued

<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 215

atggcctcta ccagatggc 19

<210> SEQ ID NO 216
<211> LENGTH: 19
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 216

atggcctctc ccagatggc 19

<210> SEQ ID NO 217
<211> LENGTH: 19
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 217

atggcctctg ccagatggc 19

<210> SEQ ID NO 218
<211> LENGTH: 19
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 218

atggcctctt ccagatggc 19

<210> SEQ ID NO 219
<211> LENGTH: 19
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 219

cagatggctt tgaaaaagg 19

<210> SEQ ID NO 220
<211> LENGTH: 19
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 220

cagatggctt tgcaaaagg 19

<210> SEQ ID NO 221
<211> LENGTH: 19
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 221

cagatggctt tggaaaagg 19

<210> SEQ ID NO 222
<211> LENGTH: 19
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 222

cagatggctt tgtaaaagg 19

<210> SEQ ID NO 223

A000596

US 6,951,721 B2

99

100

-continued

<211> LENGTH: 19
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 223

gatccaaaca ggcccccttt 19

<210> SEQ ID NO 224
<211> LENGTH: 19
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 224

gatccaacca ggcccccttt 19

<210> SEQ ID NO 225
<211> LENGTH: 19
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 225

gatccaagca ggcccccttt 19

<210> SEQ ID NO 226
<211> LENGTH: 19
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 226

gatccaatca ggcccccttt 19

<210> SEQ ID NO 227
<211> LENGTH: 19
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 227

ccctgtaaaa actggatca 19

<210> SEQ ID NO 228
<211> LENGTH: 19
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 228

ccctgtaaac actggatca 19

<210> SEQ ID NO 229
<211> LENGTH: 19
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 229

ccctgtaaag actggatca 19

<210> SEQ ID NO 230
<211> LENGTH: 19
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 230

ccctgtaaat actggatca 19

A000597

US 6,951,721 B2

101

102

-continued

```

<210> SEQ ID NO 231
<211> LENGTH: 28
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 231

tcatgctgga tccccacttt tcctcttg                28

<210> SEQ ID NO 232
<211> LENGTH: 28
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 232

ggtggcctgc cttccaatg gatccact                28

<210> SEQ ID NO 233
<211> LENGTH: 30
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 233

aattcatggg actgactttc tgctcttgtc                30

<210> SEQ ID NO 234
<211> LENGTH: 26
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 234

tccaggtccc agccaaccc ttgtcc                    26

<210> SEQ ID NO 235
<211> LENGTH: 30
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 235

gtcctctgac tgctcttttc acccatctac                30

<210> SEQ ID NO 236
<211> LENGTH: 26
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 236

gggatacggc caggcattga agtctc                    26

<210> SEQ ID NO 237
<211> LENGTH: 28
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 237

cttgtgccct gactttcaac tctgtctc                28

<210> SEQ ID NO 238
<211> LENGTH: 27
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 238

tgggcaacca gccctgtcgt ctctcca                27

```

A000598

US 6,951,721 B2

103

104

-continued

<210> SEQ ID NO 239
 <211> LENGTH: 29
 <212> TYPE: DNA
 <213> ORGANISM: Homo sapiens

<400> SEQUENCE: 239

ccaggcctct gattcctcac tgattgctc

29

<210> SEQ ID NO 240
 <211> LENGTH: 27
 <212> TYPE: DNA
 <213> ORGANISM: Homo sapiens

<400> SEQUENCE: 240

gccactgaca accaccctta acccctc

27

<210> SEQ ID NO 241
 <211> LENGTH: 27
 <212> TYPE: DNA
 <213> ORGANISM: Homo sapiens

<400> SEQUENCE: 241

gcctcatctt gggcctgtgt tatctcc

27

<210> SEQ ID NO 242
 <211> LENGTH: 28
 <212> TYPE: DNA
 <213> ORGANISM: Homo sapiens

<400> SEQUENCE: 242

ggccagtgtg cagggtggca agtggtc

28

<210> SEQ ID NO 243
 <211> LENGTH: 30
 <212> TYPE: DNA
 <213> ORGANISM: Homo sapiens

<400> SEQUENCE: 243

gtaggacctg atttccttac tgctcttgc

30

<210> SEQ ID NO 244
 <211> LENGTH: 29
 <212> TYPE: DNA
 <213> ORGANISM: Homo sapiens

<400> SEQUENCE: 244

ataactgcac ccttggtctc ctccaccgc

29

<210> SEQ ID NO 245
 <211> LENGTH: 30
 <212> TYPE: DNA
 <213> ORGANISM: Homo sapiens

<400> SEQUENCE: 245

cacttttata acctttcctt gcctctttcc

30

<210> SEQ ID NO 246
 <211> LENGTH: 30
 <212> TYPE: DNA
 <213> ORGANISM: Homo sapiens

<400> SEQUENCE: 246

aactttccac ttgataagag gtccaagac

30

A000599

US 6,951,721 B2

105

106

-continued

<210> SEQ ID NO 247
<211> LENGTH: 29
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 247

acttacttct cccctctctc tgttgctgc

29

<210> SEQ ID NO 248
<211> LENGTH: 31
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 248

atggaatcct atggctttcc aacctaggaa g

31

<210> SEQ ID NO 249
<211> LENGTH: 29
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 249

catctctcct cctgcttct gtctctac

29

<210> SEQ ID NO 250
<211> LENGTH: 30
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 250

ctgacgcaca cctattgcaa gcaagggttc

30

<210> SEQ ID NO 251
<211> LENGTH: 19
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 251

gaagcgtctc atgctggat

19

<210> SEQ ID NO 252
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 252

ggggactgta gatgggtgaa

20

<210> SEQ ID NO 253
<211> LENGTH: 19
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 253

ttttcagacc tatggaac

19

<210> SEQ ID NO 254
<211> LENGTH: 19
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 254

A000600

US 6,951,721 B2

107

108

-continued

tattcagatc	tattggaac	19
<210> SEQ ID NO 255		
<211> LENGTH: 18		
<212> TYPE: DNA		
<213> ORGANISM: Homo sapiens		
<400> SEQUENCE: 255		
cccttgccgt	cccaagca	18
<210> SEQ ID NO 256		
<211> LENGTH: 18		
<212> TYPE: DNA		
<213> ORGANISM: Homo sapiens		
<400> SEQUENCE: 256		
cccttgccat	cccaagca	18
<210> SEQ ID NO 257		
<211> LENGTH: 18		
<212> TYPE: DNA		
<213> ORGANISM: Homo sapiens		
<400> SEQUENCE: 257		
ctgtccccgg	acgatatt	18
<210> SEQ ID NO 258		
<211> LENGTH: 18		
<212> TYPE: DNA		
<213> ORGANISM: Homo sapiens		
<400> SEQUENCE: 258		
ctgtccccag	acgatatt	18
<210> SEQ ID NO 259		
<211> LENGTH: 18		
<212> TYPE: DNA		
<213> ORGANISM: Homo sapiens		
<400> SEQUENCE: 259		
gctcccccg	tggccct	18
<210> SEQ ID NO 260		
<211> LENGTH: 18		
<212> TYPE: DNA		
<213> ORGANISM: Homo sapiens		
<400> SEQUENCE: 260		
gctccccgg	tggccct	18
<210> SEQ ID NO 261		
<211> LENGTH: 18		
<212> TYPE: DNA		
<213> ORGANISM: Homo sapiens		
<400> SEQUENCE: 261		
acttttcgac	atagtgtg	18
<210> SEQ ID NO 262		
<211> LENGTH: 18		
<212> TYPE: DNA		
<213> ORGANISM: Homo sapiens		
<400> SEQUENCE: 262		

A000601

US 6,951,721 B2

109

110

-continued

acttttcggc atagtgtg 18

<210> SEQ ID NO 263
 <211> LENGTH: 5711
 <212> TYPE: DNA
 <213> ORGANISM: Homo sapiens
 <220> FEATURE:
 <221> NAME/KEY: CDS
 <222> LOCATION: (120)..(5711)
 <223> OTHER INFORMATION: omil sequences

<400> SEQUENCE: 263

agctcgctga gacttcctgg accccgcacc aggctgtggg gtttctcaga taactgggcc 60

cctgcgctca ggaggccttc accctctgct ctgggtaaag ttcatggaa cagaaagaa 119

atg gat tta tct gct ctt cgc gtt gaa gaa gta caa aat gtc att aat 167
 Met Asp Leu Ser Ala Leu Arg Val Glu Glu Val Gln Asn Val Ile Asn
 1 5 10 15

gct atg cag aaa atc tta gag tgt ccc atc tgt ctg gag ttg atc aag 215
 Ala Met Gln Lys Ile Leu Glu Cys Pro Ile Cys Leu Glu Leu Ile Lys
 20 25 30

gaa cct gtc tcc aca aag tgt gac cac ata ttt tgc aaa ttt tgc atg 263
 Glu Pro Val Ser Thr Lys Cys Asp His Ile Phe Cys Lys Phe Cys Met
 35 40 45

ctg aaa ctt ctc aac cag aag aaa ggg cct tca cag tgt cct tta tgt 311
 Leu Lys Leu Leu Asn Gln Lys Lys Gly Pro Ser Gln Cys Pro Leu Cys
 50 55 60

aag aat gat ata acc aaa agg agc cta caa gaa agt acg aga ttt agt 359
 Lys Asn Asp Ile Thr Lys Arg Ser Leu Gln Glu Ser Thr Arg Phe Ser
 65 70 75 80

caa ctt gtt gaa gag cta ttg aaa atc att tgt gct ttt cag ctt gac 407
 Gln Leu Val Glu Glu Leu Leu Lys Ile Ile Cys Ala Phe Gln Leu Asp
 85 90 95

aca ggt ttg gag tat gca aac agc tat aat ttt gca aaa aag gaa aat 455
 Thr Gly Leu Glu Tyr Ala Asn Ser Tyr Asn Phe Ala Lys Lys Glu Asn
 100 105 110

aac tct cct gaa cat cta aaa gat gaa gtt tct atc atc caa agt atg 503
 Asn Ser Pro Glu His Leu Lys Asp Glu Val Ser Ile Ile Gln Ser Met
 115 120 125

ggc tac aga aac cgt gcc aaa aga ctt cta cag agt gaa ccc gaa aat 551
 Gly Tyr Arg Asn Arg Ala Lys Arg Leu Leu Gln Ser Glu Pro Glu Asn
 130 135 140

cct tcc ttg cag gaa acc agt ctc agt gtc caa ctc tct aac ctt gga 599
 Pro Ser Leu Gln Glu Thr Ser Leu Ser Val Gln Leu Ser Asn Leu Gly
 145 150 155 160

act gtg aga act ctg agg aca aag cag cgg ata caa cct caa aag acg 647
 Thr Val Arg Thr Leu Arg Thr Lys Gln Arg Ile Gln Pro Gln Lys Thr
 165 170 175

tct gtc tac att gaa ttg gga tct gat tct tct gaa gat acc gtt aat 695
 Ser Val Tyr Ile Glu Leu Gly Ser Asp Ser Ser Glu Asp Thr Val Asn
 180 185 190

aag gca act tat tgc agt gtg gga gat caa gaa ttg tta caa atc acc 743
 Lys Ala Thr Tyr Cys Ser Val Gly Asp Gln Glu Leu Leu Gln Ile Thr
 195 200 205

cct caa gga acc agg gat gaa atc agt ttg gat tct gca aaa aag gct 791
 Pro Gln Gly Thr Arg Asp Glu Ile Ser Leu Asp Ser Ala Lys Lys Ala
 210 215 220

gct tgt gaa ttt tct gag acg gat gta aca aat act gaa cat cat caa 839
 Ala Cys Glu Phe Ser Glu Thr Asp Val Thr Asn Thr Glu His His Gln
 225 230 235 240

A000602

US 6,951,721 B2

111

112

-continued

ccc agt aat aat gat ttg aac acc act gag aag cgt gca gct gag agg Pro Ser Asn Asn Asp Leu Asn Thr Thr Glu Lys Arg Ala Ala Glu Arg 245 250 255	887
cat cca gaa aag tat cag ggt agt tct gtt tca aac ttg cat gtg gag His Pro Glu Lys Tyr Gln Gly Ser Ser Val Ser Asn Leu His Val Glu 260 265 270	935
cca tgt ggc aca aat act cat gcc agc tca tta cag cat gag aac agc Pro Cys Gly Thr Asn Thr His Ala Ser Ser Leu Gln His Glu Asn Ser 275 280 285	983
agt tta tta ctc act aaa gac aga atg aat gta gaa aag gct gaa ttc Ser Leu Leu Leu Thr Lys Asp Arg Met Asn Val Glu Lys Ala Glu Phe 290 295 300	1031
tgt aat aaa agc aaa cag cct gcc tta gca agg agc caa cat aac aga Cys Asn Lys Ser Lys Gln Pro Gly Leu Ala Arg Ser Gln His Asn Arg 305 310 315 320	1079
tgg gct gga agt aag gaa aca tgt aat gat agg cgg act ccc agc aca Trp Ala Gly Ser Lys Glu Thr Cys Asn Asp Arg Arg Thr Pro Ser Thr 325 330 335	1127
gaa aaa aag gta gat ctg aat gct gat ccc ctg tgt gag aga aaa gaa Glu Lys Lys Val Asp Leu Asn Ala Asp Pro Leu Cys Glu Arg Lys Glu 340 345 350	1175
tgg aat aag cag aaa ctg cca tgc tca gag aat cct aga gat act gaa Trp Asn Lys Gln Lys Leu Pro Cys Ser Glu Asn Pro Arg Asp Thr Glu 355 360 365	1223
gat gtt cct tgg ata aca cta aat agc agc att cag aaa gtt aat gag Asp Val Pro Trp Ile Thr Leu Asn Ser Ser Ile Gln Lys Val Asn Glu 370 375 380	1271
tgg ttt tcc aga agt gat gaa ctg tta ggt tct gat gac tca cat gat Trp Phe Ser Arg Ser Asp Glu Leu Leu Gly Ser Asp Asp Ser His Asp 385 390 395 400	1319
ggg gag tct gaa tca aat gcc aaa gta gct gat gta ttg gac gtt cta Gly Glu Ser Glu Ser Asn Ala Lys Val Ala Asp Val Leu Asp Val Leu 405 410 415	1367
aat gag gta gat gaa tat tct ggt tct tca gag aaa ata gac tta ctg Asn Glu Val Asp Glu Tyr Ser Gly Ser Ser Glu Lys Ile Asp Leu Leu 420 425 430	1415
gcc agt gat cct cat gag gct tta ata tgt aaa agt gaa aga gtt cac Ala Ser Asp Pro His Glu Ala Leu Ile Cys Lys Ser Glu Arg Val His 435 440 445	1463
tcc aaa tca gta gag agt aat att gaa gac aaa ata ttt ggg aaa acc Ser Lys Ser Val Glu Ser Asn Ile Glu Asp Lys Ile Phe Gly Lys Thr 450 455 460	1511
tat cgg aag aag gca agc ctc ccc aac tta agc cat gta act gaa aat Tyr Arg Lys Lys Ala Ser Leu Pro Asn Leu Ser His Val Thr Glu Asn 465 470 475 480	1559
cta att ata gga gca ttt gtt act gag cca cag ata ata caa gag cgt Leu Ile Ile Gly Ala Phe Val Thr Glu Pro Gln Ile Ile Gln Glu Arg 485 490 495	1607
ccc ctc aca aat aaa tta aag cgt aaa agg aga cct aca tca ggc ctt Pro Leu Thr Asn Lys Leu Lys Arg Lys Arg Arg Pro Thr Ser Gly Leu 500 505 510	1655
cat cct gag gat ttt atc aag aaa gca gat ttg gca gtt caa aag act His Pro Glu Asp Phe Ile Lys Lys Ala Asp Leu Ala Val Gln Lys Thr 515 520 525	1703
cct gaa atg ata aat cag gga act aac caa acg gag cag aat ggt caa Pro Glu Met Ile Asn Gln Gly Thr Asn Gln Thr Glu Gln Asn Gly Gln 530 535 540	1751
gtg atg aat att act aat agt ggt cat gag aat aaa aca aaa ggt gat Val Met Asn Ile Thr Asn Ser Gly His Glu Asn Lys Thr Lys Gly Asp 545 550 555 560	1799

A000603

US 6,951,721 B2

113

114

-continued

tct att cag aat gag aaa aat cct aac cca ata gaa tca ctc gaa aaa Ser Ile Gln Asn Glu Lys Asn Pro Asn Pro Ile Glu Ser Leu Glu Lys 565 570 575	1847
gaa tct gct ttc aaa acg aaa gct gaa cct ata agc agc agt ata agc Glu Ser Ala Phe Lys Thr Lys Ala Glu Pro Ile Ser Ser Ser Ile Ser 580 585 590	1895
aat atg gaa ctc gaa tta aat atc cac aat tca aaa gca cct aaa aag Asn Met Glu Leu Glu Leu Asn Ile His Asn Ser Lys Ala Pro Lys Lys 595 600 605	1943
aat agg ctg agg agg aag tct tct acc agg cat att cat gcg ctt gaa Asn Arg Leu Arg Arg Lys Ser Ser Thr Arg His Ile His Ala Leu Glu 610 615 620	1991
cta gta gtc agt aga aat cta agc cca cct aat tgt act gaa ttg caa Leu Val Val Ser Arg Asn Leu Ser Pro Pro Asn Cys Thr Glu Leu Gln 625 630 635 640	2039
att gat agt tgt tct agc agt gaa gag ata aag aaa aaa aag tac aac Ile Asp Ser Cys Ser Ser Ser Glu Glu Ile Lys Lys Lys Lys Tyr Asn 645 650 655	2087
caa atg cca gtc agg cac agc aga aac cta caa ctc atg gaa ggt aaa Gln Met Pro Val Arg His Ser Arg Asn Leu Gln Leu Met Glu Gly Lys 660 665 670	2135
gaa cct gca act gga gcc aag aag agt aac aag cca aat gaa cag aca Glu Pro Ala Thr Gly Ala Lys Lys Ser Asn Lys Pro Asn Glu Gln Thr 675 680 685	2183
agt aaa aga cat gac agt gat act ttc cca gag ctg aag tta aca aat Ser Lys Arg His Asp Ser Asp Thr Phe Pro Glu Leu Lys Leu Thr Asn 690 695 700	2231
gca cct ggt tct ttt act aag tgt tca aat acc agt gaa ctt aaa gaa Ala Pro Gly Ser Phe Thr Lys Cys Ser Asn Thr Ser Glu Leu Lys Glu 705 710 715 720	2279
ttt gtc aat cct agc ctt cca aga gaa gaa aaa gaa gag aaa cta gaa Phe Val Asn Pro Ser Leu Pro Arg Glu Glu Lys Glu Glu Lys Leu Glu 725 730 735	2327
aca gtt aaa gtg tct aat aat gct gaa gac ccc aaa gat ctc atg tta Thr Val Lys Val Ser Asn Asn Ala Glu Asp Pro Lys Asp Leu Met Leu 740 745 750	2375
agt gga gaa agg gtt ttg caa act gaa aga tct gta gag agt agc agt Ser Gly Glu Arg Val Leu Gln Thr Glu Arg Ser Val Glu Ser Ser Ser 755 760 765	2423
att tca ctg gta cct ggt act gat tat ggc act cag gaa agt atc tcg Ile Ser Leu Val Pro Gly Thr Asp Tyr Gly Thr Gln Glu Ser Ile Ser 770 775 780	2471
tta ctg gaa gtt agc act cta ggg aag gca aaa aca gaa cca aat aaa Leu Leu Glu Val Ser Thr Leu Gly Lys Ala Lys Thr Glu Pro Asn Lys 785 790 795 800	2519
tgt gtg agt cag tgt gca gca ttt gaa aac ccc aag gga cta att cat Cys Val Ser Gln Cys Ala Ala Phe Glu Asn Pro Lys Gly Leu Ile His 805 810 815	2567
ggt tgt tcc aaa gat aat aga aat gac aca gaa ggc ttt aag tat cca Gly Cys Ser Lys Asp Asn Arg Asn Asp Thr Glu Gly Phe Lys Tyr Pro 820 825 830	2615
ttg gga cat gaa gtt aac cac agt cgg gaa aca agc ata gaa atg gaa Leu Gly His Glu Val Asn His Ser Arg Glu Thr Ser Ile Glu Met Glu 835 840 845	2663
gaa agt gaa ctt gat gct cag tat ttg cag aat aca ttc aag gtt tca Glu Ser Glu Leu Asp Ala Gln Tyr Leu Gln Asn Thr Phe Lys Val Ser 850 855 860	2711
aag cgc cag tca ttt gct ctg ttt tca aat cca gga aat gca gaa gag Lys Arg Gln Ser Phe Ala Leu Phe Ser Asn Pro Gly Asn Ala Glu Glu	2759

A000604

US 6,951,721 B2

115

116

-continued

865	870	875	880	
gaa tgt gca aca ttc tct gcc cac tct ggg tcc tta aag aaa caa agt				2807
Glu Cys Ala Thr Phe Ser Ala His Ser Gly Ser Leu Lys Lys Gln Ser				
885 890 895				
cca aaa gtc act ttt gaa tgt gaa caa aag gaa gaa aat caa gga aag				2855
Pro Lys Val Thr Phe Glu Cys Glu Gln Lys Glu Glu Asn Gln Gly Lys				
900 905 910				
aat gag tct aat atc aag cct gta cag aca gtt aat atc act gca ggc				2903
Asn Glu Ser Asn Ile Lys Pro Val Gln Thr Val Asn Ile Thr Ala Gly				
915 920 925				
ttt cct gtg gtt ggt cag aaa gat aag cca gtt gat aat gcc aaa tgt				2951
Phe Pro Val Val Gly Gln Lys Asp Lys Pro Val Asp Asn Ala Lys Cys				
930 935 940				
agt atc aaa gga ggc tct agg ttt tgt cta tca tct cag ttc aga ggc				2999
Ser Ile Lys Gly Gly Ser Arg Phe Cys Leu Ser Ser Gln Phe Arg Gly				
945 950 955 960				
aac gaa act gga ctc att act cca aat aaa cat gga ctt tta caa aac				3047
Asn Glu Thr Gly Leu Ile Thr Pro Asn Lys His Gly Leu Leu Gln Asn				
965 970 975				
cca tat cgt ata cca cca ctt ttt ccc atc aag tca ttt gtt aaa act				3095
Pro Tyr Arg Ile Pro Pro Leu Phe Pro Ile Lys Ser Phe Val Lys Thr				
980 985 990				
aaa tgt aag aaa aat ctg cta gag gaa aac ttt gag gaa cat tca atg				3143
Lys Cys Lys Lys Asn Leu Leu Glu Glu Asn Phe Glu Glu His Ser Met				
995 1000 1005				
tca cct gaa aga gaa atg gga aat gag aac att cca agt aca gtg				3188
Ser Pro Glu Arg Glu Met Gly Asn Glu Asn Ile Pro Ser Thr Val				
1010 1015 1020				
agc aca att agc cgt aat aac att aga gaa aat gtt ttt aaa gga				3233
Ser Thr Ile Ser Arg Asn Asn Ile Arg Glu Asn Val Phe Lys Gly				
1025 1030 1035				
gcc agc tca agc aat att aat gaa gta ggt tcc agt act aat gaa				3278
Ala Ser Ser Ser Asn Ile Asn Glu Val Gly Ser Ser Thr Asn Glu				
1040 1045 1050				
gtg ggc tcc agt att aat gaa ata ggt tcc agt gat gaa aac att				3323
Val Gly Ser Ser Ile Asn Glu Ile Gly Ser Ser Asp Glu Asn Ile				
1055 1060 1065				
caa gca gaa cta ggt aga aac aga ggg cca aaa ttg aat gct atg				3368
Gln Ala Glu Leu Gly Arg Asn Arg Gly Pro Lys Leu Asn Ala Met				
1070 1075 1080				
ctt aga tta ggg gtt ttg caa cct gag gtc tat aaa caa agt ctt				3413
Leu Arg Leu Gly Val Leu Gln Pro Glu Val Tyr Lys Gln Ser Leu				
1085 1090 1095				
cct gga agt aat tgt aag cat cct gaa ata aaa aag caa gaa tat				3458
Pro Gly Ser Asn Cys Lys His Pro Glu Ile Lys Lys Gln Glu Tyr				
1100 1105 1110				
gaa gaa gta gtt cag act gtt aat aca gat ttc tct cca tat ctg				3503
Glu Glu Val Val Gln Thr Val Asn Thr Asp Phe Ser Pro Tyr Leu				
1115 1120 1125				
att tca gat aac tta gaa cag cct atg gga agt agt cat gca tct				3548
Ile Ser Asp Asn Leu Glu Gln Pro Met Gly Ser Ser His Ala Ser				
1130 1135 1140				
cag gtt tgt tct gag aca cct gat gac ctg tta gat gat ggt gaa				3593
Gln Val Cys Ser Glu Thr Pro Asp Asp Leu Leu Asp Asp Gly Glu				
1145 1150 1155				
ata aag gaa gat act agt ttt gct gaa aat gac att aag gaa agt				3638
Ile Lys Glu Asp Thr Ser Phe Ala Glu Asn Asp Ile Lys Glu Ser				
1160 1165 1170				
tct gct gtt ttt agc aaa agc gtc cag aga gga gag ctt agc agg				3683

A000605

US 6,951,721 B2

117

118

-continued

Ser	Ala	Val	Phe	Ser	Lys	Ser	Val	Gln	Arg	Gly	Glu	Leu	Ser	Arg	
1175						1180					1185				
agt	cct	agc	cct	ttc	acc	cat	aca	cat	ttg	gct	cag	ggg	tac	cga	3728
Ser	Pro	Ser	Pro	Phe	Thr	His	Thr	His	Leu	Ala	Gln	Gly	Tyr	Arg	
1190						1195					1200				
aga	ggg	gcc	aag	aaa	tta	gag	tcc	tca	gaa	gag	aac	tta	tct	agt	3773
Arg	Gly	Ala	Lys	Lys	Leu	Glu	Ser	Ser	Glu	Glu	Asn	Leu	Ser	Ser	
1205						1210					1215				
gag	gat	gaa	gag	ctt	ccc	tgc	ttc	caa	cac	ttg	tta	ttt	ggg	aaa	3818
Glu	Asp	Glu	Glu	Leu	Pro	Cys	Phe	Gln	His	Leu	Leu	Phe	Gly	Lys	
1220						1225					1230				
gta	aac	aat	ata	cct	tct	cag	tct	act	agg	cat	agc	acc	gtt	gct	3863
Val	Asn	Asn	Ile	Pro	Ser	Gln	Ser	Thr	Arg	His	Ser	Thr	Val	Ala	
1235						1240					1245				
acc	gag	tgt	ctg	tct	aag	aac	aca	gag	gag	aat	tta	tta	tca	ttg	3908
Thr	Glu	Cys	Leu	Ser	Lys	Asn	Thr	Glu	Glu	Asn	Leu	Leu	Ser	Leu	
1250						1255					1260				
aag	aat	agc	tta	aat	gac	tgc	agt	aac	cag	gta	ata	ttg	gca	aag	3953
Lys	Asn	Ser	Leu	Asn	Asp	Cys	Ser	Asn	Gln	Val	Ile	Leu	Ala	Lys	
1265						1270					1275				
gca	tct	cag	gaa	cat	cac	ctt	agt	gag	gaa	aca	aaa	tgt	tct	gct	3998
Ala	Ser	Gln	Glu	His	His	Leu	Ser	Glu	Glu	Thr	Lys	Cys	Ser	Ala	
1280						1285					1290				
agc	ttg	ttt	tct	tca	cag	tgc	agt	gaa	ttg	gaa	gac	ttg	act	gca	4043
Ser	Leu	Phe	Ser	Ser	Gln	Cys	Ser	Glu	Leu	Glu	Asp	Leu	Thr	Ala	
1295						1300					1305				
aat	aca	aac	acc	cag	gat	cct	ttc	ttg	att	ggg	tct	tcc	aaa	caa	4088
Asn	Thr	Asn	Thr	Gln	Asp	Pro	Phe	Leu	Ile	Gly	Ser	Ser	Lys	Gln	
1310						1315					1320				
atg	agg	cat	cag	tct	gaa	agc	cag	gga	gtt	ggg	ctg	agt	gac	aag	4133
Met	Arg	His	Gln	Ser	Glu	Ser	Gln	Gly	Val	Gly	Leu	Ser	Asp	Lys	
1325						1330					1335				
gaa	ttg	gtt	tca	gat	gat	gaa	gaa	aga	gga	acg	ggc	ttg	gaa	gaa	4178
Glu	Leu	Val	Ser	Asp	Asp	Glu	Glu	Arg	Gly	Thr	Gly	Leu	Glu	Glu	
1340						1345					1350				
aat	aat	caa	gaa	gag	caa	agc	atg	gat	tca	aac	tta	ggg	gaa	gca	4223
Asn	Asn	Gln	Glu	Glu	Gln	Ser	Met	Asp	Ser	Asn	Leu	Gly	Glu	Ala	
1355						1360					1365				
gca	tct	ggg	tgt	gag	agt	gaa	aca	agc	gtc	tct	gaa	gac	tgc	tca	4268
Ala	Ser	Gly	Cys	Glu	Ser	Glu	Thr	Ser	Val	Ser	Glu	Asp	Cys	Ser	
1370						1375					1380				
ggg	cta	tcc	tct	cag	agt	gac	att	tta	acc	act	cag	cag	agg	gat	4313
Gly	Leu	Ser	Ser	Gln	Ser	Asp	Ile	Leu	Thr	Thr	Gln	Gln	Arg	Asp	
1385						1390					1395				
acc	atg	caa	cat	aac	ctg	ata	aag	ctc	cag	cag	gaa	atg	gct	gaa	4358
Thr	Met	Gln	His	Asn	Leu	Ile	Lys	Leu	Gln	Gln	Glu	Met	Ala	Glu	
1400						1405					1410				
cta	gaa	gct	gtg	tta	gaa	cag	cat	ggg	agc	cag	cct	tct	aac	agc	4403
Leu	Glu	Ala	Val	Leu	Glu	Gln	His	Gly	Ser	Gln	Pro	Ser	Asn	Ser	
1415						1420					1425				
tac	cct	tcc	atc	ata	agt	gac	tcc	tct	gcc	ctt	gag	gac	ctg	cga	4448
Tyr	Pro	Ser	Ile	Ile	Ser	Asp	Ser	Ser	Ala	Leu	Glu	Asp	Leu	Arg	
1430						1435					1440				
aat	cca	gaa	caa	agc	aca	tca	gaa	aaa	gca	gta	tta	act	tca	cag	4493
Asn	Pro	Glu	Gln	Ser	Thr	Ser	Glu	Lys	Ala	Val	Leu	Thr	Ser	Gln	
1445						1450					1455				
aaa	agt	agt	gaa	tac	cct	ata	agc	cag	aat	cca	gaa	ggc	ctt	tct	4538
Lys	Ser	Ser	Glu	Tyr	Pro	Ile	Ser	Gln	Asn	Pro	Glu	Gly	Leu	Ser	
1460						1465					1470				

A000606

US 6,951,721 B2

119

120

-continued

gct gac aag ttt gag gtg tct gca gat agt tct acc agt aaa aat 4583 Ala Asp Lys Phe Glu Val Ser Ala Asp Ser Ser Thr Ser Lys Asn 1475 1480 1485
aaa gaa cca gga gtg gaa agg tca tcc cct tct aaa tgc cca tca 4628 Lys Glu Pro Gly Val Glu Arg Ser Ser Pro Ser Lys Cys Pro Ser 1490 1495 1500
tta gat gat agg tgg tac atg cac agt tgc tct ggg agt ctt cag 4673 Leu Asp Asp Arg Trp Tyr Met His Ser Cys Ser Gly Ser Leu Gln 1505 1510 1515
aat aga aac tac cca tct caa gag gag ctc att aag gtt gtt gat 4718 Asn Arg Asn Tyr Pro Ser Gln Glu Glu Leu Ile Lys Val Val Asp 1520 1525 1530
gtg gag gag caa cag ctg gaa gag tct ggg cca cac gat ttg acg 4763 Val Glu Glu Gln Gln Leu Glu Glu Ser Gly Pro His Asp Leu Thr 1535 1540 1545
gaa aca tct tac ttg cca agg caa gat cta gag gga acc cct tac 4808 Glu Thr Ser Tyr Leu Pro Arg Gln Asp Leu Glu Gly Thr Pro Tyr 1550 1555 1560
ctg gaa tct gga atc agc ctc ttc tct gat gac cct gaa tct gat 4853 Leu Glu Ser Gly Ile Ser Leu Phe Ser Asp Asp Pro Glu Ser Asp 1565 1570 1575
cct tct gaa gac aga gcc cca gag tca gct cgt gtt ggc aac ata 4898 Pro Ser Glu Asp Arg Ala Pro Glu Ser Ala Arg Val Gly Asn Ile 1580 1585 1590
cca tct tca acc tct gca ttg aaa gtt ccc caa ttg aaa gtt gca 4943 Pro Ser Ser Thr Ser Ala Leu Lys Val Pro Gln Leu Lys Val Ala 1595 1600 1605
gaa tct gcc cag ggt cca gct gct gct cat act act gat act gct 4988 Glu Ser Ala Gln Gly Pro Ala Ala Ala His Thr Thr Asp Thr Ala 1610 1615 1620
ggg tat aat gca atg gaa gaa agt gtg agc agg gag aag cca gaa 5033 Gly Tyr Asn Ala Met Glu Glu Ser Val Ser Arg Glu Lys Pro Glu 1625 1630 1635
ttg aca gct tca aca gaa agg gtc aac aaa aga atg tcc atg gtg 5078 Leu Thr Ala Ser Thr Glu Arg Val Asn Lys Arg Met Ser Met Val 1640 1645 1650
gtg tct ggc ctg acc cca gaa gaa ttt atg ctc gtg tac aag ttt 5123 Val Ser Gly Leu Thr Pro Glu Glu Phe Met Leu Val Tyr Lys Phe 1655 1660 1665
gcc aga aaa cac cac atc act tta act aat cta att act gaa gag 5168 Ala Arg Lys His His Ile Thr Leu Thr Asn Leu Ile Thr Glu Glu 1670 1675 1680
act act cat gtt gtt atg aaa aca gat gct gag ttt gtg tgt gaa 5213 Thr Thr His Val Val Met Lys Thr Asp Ala Glu Phe Val Cys Glu 1685 1690 1695
cgg aca ctg aaa tat ttt cta gga att gcg gga gga aaa tgg gta 5258 Arg Thr Leu Lys Tyr Phe Leu Gly Ile Ala Gly Gly Lys Trp Val 1700 1705 1710
gtt agc tat ttc tgg gtg acc cag tct att aaa gaa aga aaa atg 5303 Val Ser Tyr Phe Trp Val Thr Gln Ser Ile Lys Glu Arg Lys Met 1715 1720 1725
ctg aat gag cat gat ttt gaa gtc aga gga gat gtg gtc aat gga 5348 Leu Asn Glu His Asp Phe Glu Val Arg Gly Asp Val Val Asn Gly 1730 1735 1740
aga aac cac caa ggt cca aag cga gca aga gaa tcc cag gac aga 5393 Arg Asn His Gln Gly Pro Lys Arg Ala Arg Glu Ser Gln Asp Arg 1745 1750 1755
aag atc ttc agg ggg cta gaa atc tgt tgc tat ggg ccc ttc acc 5438 Lys Ile Phe Arg Gly Leu Glu Ile Cys Cys Tyr Gly Pro Phe Thr 1760 1765 1770

A000607

US 6,951,721 B2

121

122

-continued

```

aac atg ccc aca gat caa ctg gaa tgg atg gta cag ctg tgt ggt 5483
Asn Met Pro Thr Asp Gln Leu Glu Trp Met Val Gln Leu Cys Gly
1775 1780 1785

gct tct gtg gtg aag gag ctt tca tca ttc acc ctt ggc aca ggt 5528
Ala Ser Val Val Lys Glu Leu Ser Ser Phe Thr Leu Gly Thr Gly
1790 1795 1800

gtc cac cca att gtg gtt gtg cag cca gat gcc tgg aca gag gac 5573
Val His Pro Ile Val Val Val Gln Pro Asp Ala Trp Thr Glu Asp
1805 1810 1815

aat ggc ttc cat gca att ggg cag atg tgt gag gca cct gtg gtg 5618
Asn Gly Phe His Ala Ile Gly Gln Met Cys Glu Ala Pro Val Val
1820 1825 1830

acc cga gag tgg gtg ttg gac agt gta gca ctc tac cag tgc cag 5663
Thr Arg Glu Trp Val Leu Asp Ser Val Ala Leu Tyr Gln Cys Gln
1835 1840 1845

gag ctg gac acc tac ctg ata ccc cag atc ccc cac agc cac tac 5708
Glu Leu Asp Thr Tyr Leu Ile Pro Gln Ile Pro His Ser His Tyr
1850 1855 1860

tga 5711

```

```

<210> SEQ ID NO 264
<211> LENGTH: 1863
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

```

```

<400> SEQUENCE: 264

```

```

Met Asp Leu Ser Ala Leu Arg Val Glu Glu Val Gln Asn Val Ile Asn
1      5      10      15

Ala Met Gln Lys Ile Leu Glu Cys Pro Ile Cys Leu Glu Leu Ile Lys
20     25     30

Glu Pro Val Ser Thr Lys Cys Asp His Ile Phe Cys Lys Phe Cys Met
35     40     45

Leu Lys Leu Leu Asn Gln Lys Lys Gly Pro Ser Gln Cys Pro Leu Cys
50     55     60

Lys Asn Asp Ile Thr Lys Arg Ser Leu Gln Glu Ser Thr Arg Phe Ser
65     70     75     80

Gln Leu Val Glu Glu Leu Leu Lys Ile Ile Cys Ala Phe Gln Leu Asp
85     90     95

Thr Gly Leu Glu Tyr Ala Asn Ser Tyr Asn Phe Ala Lys Lys Glu Asn
100    105    110

Asn Ser Pro Glu His Leu Lys Asp Glu Val Ser Ile Ile Gln Ser Met
115    120    125

Gly Tyr Arg Asn Arg Ala Lys Arg Leu Leu Gln Ser Glu Pro Glu Asn
130    135    140

Pro Ser Leu Gln Glu Thr Ser Leu Ser Val Gln Leu Ser Asn Leu Gly
145    150    155    160

Thr Val Arg Thr Leu Arg Thr Lys Gln Arg Ile Gln Pro Gln Lys Thr
165    170    175

Ser Val Tyr Ile Glu Leu Gly Ser Asp Ser Ser Glu Asp Thr Val Asn
180    185    190

Lys Ala Thr Tyr Cys Ser Val Gly Asp Gln Glu Leu Leu Gln Ile Thr
195    200    205

Pro Gln Gly Thr Arg Asp Glu Ile Ser Leu Asp Ser Ala Lys Lys Ala
210    215    220

Ala Cys Glu Phe Ser Glu Thr Asp Val Thr Asn Thr Glu His His Gln
225    230    235    240

```

A000608

US 6,951,721 B2

123

124

-continued

Pro Ser Asn Asn Asp Leu Asn Thr Thr Glu Lys Arg Ala Ala Glu Arg
 245 250 255
 His Pro Glu Lys Tyr Gln Gly Ser Ser Val Ser Asn Leu His Val Glu
 260 265 270
 Pro Cys Gly Thr Asn Thr His Ala Ser Ser Leu Gln His Glu Asn Ser
 275 280 285
 Ser Leu Leu Leu Thr Lys Asp Arg Met Asn Val Glu Lys Ala Glu Phe
 290 295 300
 Cys Asn Lys Ser Lys Gln Pro Gly Leu Ala Arg Ser Gln His Asn Arg
 305 310 315 320
 Trp Ala Gly Ser Lys Glu Thr Cys Asn Asp Arg Arg Thr Pro Ser Thr
 325 330 335
 Glu Lys Lys Val Asp Leu Asn Ala Asp Pro Leu Cys Glu Arg Lys Glu
 340 345 350
 Trp Asn Lys Gln Lys Leu Pro Cys Ser Glu Asn Pro Arg Asp Thr Glu
 355 360 365
 Asp Val Pro Trp Ile Thr Leu Asn Ser Ser Ile Gln Lys Val Asn Glu
 370 375 380
 Trp Phe Ser Arg Ser Asp Glu Leu Leu Gly Ser Asp Asp Ser His Asp
 385 390 395 400
 Gly Glu Ser Glu Ser Asn Ala Lys Val Ala Asp Val Leu Asp Val Leu
 405 410 415
 Asn Glu Val Asp Glu Tyr Ser Gly Ser Ser Glu Lys Ile Asp Leu Leu
 420 425 430
 Ala Ser Asp Pro His Glu Ala Leu Ile Cys Lys Ser Glu Arg Val His
 435 440 445
 Ser Lys Ser Val Glu Ser Asn Ile Glu Asp Lys Ile Phe Gly Lys Thr
 450 455 460
 Tyr Arg Lys Lys Ala Ser Leu Pro Asn Leu Ser His Val Thr Glu Asn
 465 470 475 480
 Leu Ile Ile Gly Ala Phe Val Thr Glu Pro Gln Ile Ile Gln Glu Arg
 485 490 495
 Pro Leu Thr Asn Lys Leu Lys Arg Lys Arg Arg Pro Thr Ser Gly Leu
 500 505 510
 His Pro Glu Asp Phe Ile Lys Lys Ala Asp Leu Ala Val Gln Lys Thr
 515 520 525
 Pro Glu Met Ile Asn Gln Gly Thr Asn Gln Thr Glu Gln Asn Gly Gln
 530 535 540
 Val Met Asn Ile Thr Asn Ser Gly His Glu Asn Lys Thr Lys Gly Asp
 545 550 555 560
 Ser Ile Gln Asn Glu Lys Asn Pro Asn Pro Ile Glu Ser Leu Glu Lys
 565 570 575
 Glu Ser Ala Phe Lys Thr Lys Ala Glu Pro Ile Ser Ser Ser Ile Ser
 580 585 590
 Asn Met Glu Leu Glu Leu Asn Ile His Asn Ser Lys Ala Pro Lys Lys
 595 600 605
 Asn Arg Leu Arg Arg Lys Ser Ser Thr Arg His Ile His Ala Leu Glu
 610 615 620
 Leu Val Val Ser Arg Asn Leu Ser Pro Pro Asn Cys Thr Glu Leu Gln
 625 630 635 640
 Ile Asp Ser Cys Ser Ser Ser Glu Glu Ile Lys Lys Lys Lys Tyr Asn
 645 650 655

A000609

US 6,951,721 B2

125

126

-continued

Gln Met Pro Val Arg His Ser Arg Asn Leu Gln Leu Met Glu Gly Lys	660	665	670
Glu Pro Ala Thr Gly Ala Lys Lys Ser Asn Lys Pro Asn Glu Gln Thr	675	680	685
Ser Lys Arg His Asp Ser Asp Thr Phe Pro Glu Leu Lys Leu Thr Asn	690	695	700
Ala Pro Gly Ser Phe Thr Lys Cys Ser Asn Thr Ser Glu Leu Lys Glu	705	710	715
Phe Val Asn Pro Ser Leu Pro Arg Glu Glu Lys Glu Glu Lys Leu Glu	725	730	735
Thr Val Lys Val Ser Asn Asn Ala Glu Asp Pro Lys Asp Leu Met Leu	740	745	750
Ser Gly Glu Arg Val Leu Gln Thr Glu Arg Ser Val Glu Ser Ser Ser	755	760	765
Ile Ser Leu Val Pro Gly Thr Asp Tyr Gly Thr Gln Glu Ser Ile Ser	770	775	780
Leu Leu Glu Val Ser Thr Leu Gly Lys Ala Lys Thr Glu Pro Asn Lys	785	790	795
Cys Val Ser Gln Cys Ala Ala Phe Glu Asn Pro Lys Gly Leu Ile His	805	810	815
Gly Cys Ser Lys Asp Asn Arg Asn Asp Thr Glu Gly Phe Lys Tyr Pro	820	825	830
Leu Gly His Glu Val Asn His Ser Arg Glu Thr Ser Ile Glu Met Glu	835	840	845
Glu Ser Glu Leu Asp Ala Gln Tyr Leu Gln Asn Thr Phe Lys Val Ser	850	855	860
Lys Arg Gln Ser Phe Ala Leu Phe Ser Asn Pro Gly Asn Ala Glu Glu	865	870	875
Glu Cys Ala Thr Phe Ser Ala His Ser Gly Ser Leu Lys Lys Gln Ser	885	890	895
Pro Lys Val Thr Phe Glu Cys Glu Gln Lys Glu Glu Asn Gln Gly Lys	900	905	910
Asn Glu Ser Asn Ile Lys Pro Val Gln Thr Val Asn Ile Thr Ala Gly	915	920	925
Phe Pro Val Val Gly Gln Lys Asp Lys Pro Val Asp Asn Ala Lys Cys	930	935	940
Ser Ile Lys Gly Gly Ser Arg Phe Cys Leu Ser Ser Gln Phe Arg Gly	945	950	955
Asn Glu Thr Gly Leu Ile Thr Pro Asn Lys His Gly Leu Leu Gln Asn	965	970	975
Pro Tyr Arg Ile Pro Pro Leu Phe Pro Ile Lys Ser Phe Val Lys Thr	980	985	990
Lys Cys Lys Lys Asn Leu Leu Glu Glu Asn Phe Glu Glu His Ser Met	995	1000	1005
Ser Pro Glu Arg Glu Met Gly Asn Glu Asn Ile Pro Ser Thr Val	1010	1015	1020
Ser Thr Ile Ser Arg Asn Asn Ile Arg Glu Asn Val Phe Lys Gly	1025	1030	1035
Ala Ser Ser Ser Asn Ile Asn Glu Val Gly Ser Ser Thr Asn Glu	1040	1045	1050
Val Gly Ser Ser Ile Asn Glu Ile Gly Ser Ser Asp Glu Asn Ile	1055	1060	1065
Gln Ala Glu Leu Gly Arg Asn Arg Gly Pro Lys Leu Asn Ala Met			

A000610

US 6,951,721 B2

127

128

-continued

1070	1075	1080
Leu Arg Leu Gly Val Leu Gln	Pro Glu Val Tyr Lys	Gln Ser Leu
1085	1090	1095
Pro Gly Ser Asn Cys Lys His	Pro Glu Ile Lys Lys	Gln Glu Tyr
1100	1105	1110
Glu Glu Val Val Gln Thr Val	Asn Thr Asp Phe Ser	Pro Tyr Leu
1115	1120	1125
Ile Ser Asp Asn Leu Glu Gln	Pro Met Gly Ser Ser	His Ala Ser
1130	1135	1140
Gln Val Cys Ser Glu Thr Pro	Asp Asp Leu Leu Asp	Asp Gly Glu
1145	1150	1155
Ile Lys Glu Asp Thr Ser Phe	Ala Glu Asn Asp Ile	Lys Glu Ser
1160	1165	1170
Ser Ala Val Phe Ser Lys Ser	Val Gln Arg Gly Glu	Leu Ser Arg
1175	1180	1185
Ser Pro Ser Pro Phe Thr His	Thr His Leu Ala Gln	Gly Tyr Arg
1190	1195	1200
Arg Gly Ala Lys Lys Leu Glu	Ser Ser Glu Glu Asn	Leu Ser Ser
1205	1210	1215
Glu Asp Glu Glu Leu Pro Cys	Phe Gln His Leu Leu	Phe Gly Lys
1220	1225	1230
Val Asn Asn Ile Pro Ser Gln	Ser Thr Arg His Ser	Thr Val Ala
1235	1240	1245
Thr Glu Cys Leu Ser Lys Asn	Thr Glu Glu Asn Leu	Leu Ser Leu
1250	1255	1260
Lys Asn Ser Leu Asn Asp Cys	Ser Asn Gln Val Ile	Leu Ala Lys
1265	1270	1275
Ala Ser Gln Glu His His Leu	Ser Glu Glu Thr Lys	Cys Ser Ala
1280	1285	1290
Ser Leu Phe Ser Ser Gln Cys	Ser Glu Leu Glu Asp	Leu Thr Ala
1295	1300	1305
Asn Thr Asn Thr Gln Asp Pro	Phe Leu Ile Gly Ser	Ser Lys Gln
1310	1315	1320
Met Arg His Gln Ser Glu Ser	Gln Gly Val Gly Leu	Ser Asp Lys
1325	1330	1335
Glu Leu Val Ser Asp Asp Glu	Glu Arg Gly Thr Gly	Leu Glu Glu
1340	1345	1350
Asn Asn Gln Glu Glu Gln Ser	Met Asp Ser Asn Leu	Gly Glu Ala
1355	1360	1365
Ala Ser Gly Cys Glu Ser Glu	Thr Ser Val Ser Glu	Asp Cys Ser
1370	1375	1380
Gly Leu Ser Ser Gln Ser Asp	Ile Leu Thr Thr Gln	Gln Arg Asp
1385	1390	1395
Thr Met Gln His Asn Leu Ile	Lys Leu Gln Gln Glu	Met Ala Glu
1400	1405	1410
Leu Glu Ala Val Leu Glu Gln	His Gly Ser Gln Pro	Ser Asn Ser
1415	1420	1425
Tyr Pro Ser Ile Ile Ser Asp	Ser Ser Ala Leu Glu	Asp Leu Arg
1430	1435	1440
Asn Pro Glu Gln Ser Thr Ser	Glu Lys Ala Val Leu	Thr Ser Gln
1445	1450	1455
Lys Ser Ser Glu Tyr Pro Ile	Ser Gln Asn Pro Glu	Gly Leu Ser
1460	1465	1470

A000611

US 6,951,721 B2

129

130

-continued

Ala Asp	Lys Phe	Glu Val	Ser	Ala Asp	Ser Ser	Thr	Ser Lys	Asn
1475			1480			1485		
Lys Glu	Pro Gly	Val Glu	Arg	Ser Ser	Pro Ser	Lys	Cys Pro	Ser
1490			1495			1500		
Leu Asp	Asp Arg	Trp Tyr	Met	His Ser	Cys Ser	Gly	Ser Leu	Gln
1505			1510			1515		
Asn Arg	Asn Tyr	Pro Ser	Gln	Glu Glu	Leu Ile	Lys	Val Val	Asp
1520			1525			1530		
Val Glu	Glu Gln	Gln Leu	Glu	Glu Ser	Gly Pro	His	Asp Leu	Thr
1535			1540			1545		
Glu Thr	Ser Tyr	Leu Pro	Arg	Gln Asp	Leu Glu	Gly	Thr Pro	Tyr
1550			1555			1560		
Leu Glu	Ser Gly	Ile Ser	Leu	Phe Ser	Asp Asp	Pro	Glu Ser	Asp
1565			1570			1575		
Pro Ser	Glu Asp	Arg Ala	Pro	Glu Ser	Ala Arg	Val	Gly Asn	Ile
1580			1585			1590		
Pro Ser	Ser Thr	Ser Ala	Leu	Lys Val	Pro Gln	Leu	Lys Val	Ala
1595			1600			1605		
Glu Ser	Ala Gln	Gly Pro	Ala	Ala Ala	His Thr	Thr	Asp Thr	Ala
1610			1615			1620		
Gly Tyr	Asn Ala	Met Glu	Glu	Ser Val	Ser Arg	Glu	Lys Pro	Glu
1625			1630			1635		
Leu Thr	Ala Ser	Thr Glu	Arg	Val Asn	Lys Arg	Met	Ser Met	Val
1640			1645			1650		
Val Ser	Gly Leu	Thr Pro	Glu	Glu Phe	Met Leu	Val	Tyr Lys	Phe
1655			1660			1665		
Ala Arg	Lys His	His Ile	Thr	Leu Thr	Asn Leu	Ile	Thr Glu	Glu
1670			1675			1680		
Thr Thr	His Val	Val Met	Lys	Thr Asp	Ala Glu	Phe	Val Cys	Glu
1685			1690			1695		
Arg Thr	Leu Lys	Tyr Phe	Leu	Gly Ile	Ala Gly	Gly	Lys Trp	Val
1700			1705			1710		
Val Ser	Tyr Phe	Trp Val	Thr	Gln Ser	Ile Lys	Glu	Arg Lys	Met
1715			1720			1725		
Leu Asn	Glu His	Asp Phe	Glu	Val Arg	Gly Asp	Val	Val Asn	Gly
1730			1735			1740		
Arg Asn	His Gln	Gly Pro	Lys	Arg Ala	Arg Glu	Ser	Gln Asp	Arg
1745			1750			1755		
Lys Ile	Phe Arg	Gly Leu	Glu	Ile Cys	Cys Tyr	Gly	Pro Phe	Thr
1760			1765			1770		
Asn Met	Pro Thr	Asp Gln	Leu	Glu Trp	Met Val	Gln	Leu Cys	Gly
1775			1780			1785		
Ala Ser	Val Val	Lys Glu	Leu	Ser Ser	Phe Thr	Leu	Gly Thr	Gly
1790			1795			1800		
Val His	Pro Ile	Val Val	Val	Gln Pro	Asp Ala	Trp	Thr Glu	Asp
1805			1810			1815		
Asn Gly	Phe His	Ala Ile	Gly	Gln Met	Cys Glu	Ala	Pro Val	Val
1820			1825			1830		
Thr Arg	Glu Trp	Val Leu	Asp	Ser Val	Ala Leu	Tyr	Gln Cys	Gln
1835			1840			1845		
Glu Leu	Asp Thr	Tyr Leu	Ile	Pro Gln	Ile Pro	His	Ser His	Tyr
1850			1855			1860		

A000612

US 6,951,721 B2

131

132

-continued

```

<210> SEQ ID NO 265
<211> LENGTH: 5711
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (120)..(5711)
<223> OTHER INFORMATION: omi2 sequences

<400> SEQUENCE: 265

agctcgctga gacttctctgg accccgcacc aggctgtggg gtttctcaga taactgggcc      60
cctgcgcctca ggaggccttc accctctgct ctgggtaaag ttcattggaa cagaaagaa      119
atg gat tta tct gct ctt cgc gtt gaa gaa gta caa aat gtc att aat      167
Met Asp Leu Ser Ala Leu Arg Val Glu Glu Val Gln Asn Val Ile Asn
1      5      10      15
gct atg cag aaa atc tta gag tgt ccc atc tgt ctg gag ttg atc aag      215
Ala Met Gln Lys Ile Leu Glu Cys Pro Ile Cys Leu Glu Leu Ile Lys
20      25      30
gaa cct gtc tcc aca aag tgt gac cac ata ttt tgc aaa ttt tgc atg      263
Glu Pro Val Ser Thr Lys Cys Asp His Ile Phe Cys Lys Phe Cys Met
35      40      45
ctg aaa ctt ctc aac cag aag aaa ggg cct tca cag tgt cct tta tgt      311
Leu Lys Leu Leu Asn Gln Lys Lys Gly Pro Ser Gln Cys Pro Leu Cys
50      55      60
aag aat gat ata acc aaa agg agc cta caa gaa agt acg aga ttt agt      359
Lys Asn Asp Ile Thr Lys Arg Ser Leu Gln Glu Ser Thr Arg Phe Ser
65      70      75      80
caa ctt gtt gaa gag cta ttg aaa atc att tgt gct ttt cag ctt gac      407
Gln Leu Val Glu Glu Leu Leu Lys Ile Ile Cys Ala Phe Gln Leu Asp
85      90      95
aca ggt ttg gag tat gca aac agc tat aat ttt gca aaa aag gaa aat      455
Thr Gly Leu Glu Tyr Ala Asn Ser Tyr Asn Phe Ala Lys Lys Glu Asn
100      105      110
aac tct cct gaa cat cta aaa gat gaa gtt tct atc atc caa agt atg      503
Asn Ser Pro Glu His Leu Lys Asp Glu Val Ser Ile Ile Gln Ser Met
115      120      125
ggc tac aga aac cgt gcc aaa aga ctt cta cag agt gaa ccc gaa aat      551
Gly Tyr Arg Asn Arg Ala Lys Arg Leu Leu Gln Ser Glu Pro Glu Asn
130      135      140
cct tcc ttg cag gaa acc agt ctc agt gtc caa ctc tct aac ctt gga      599
Pro Ser Leu Gln Glu Thr Ser Leu Ser Val Gln Leu Ser Asn Leu Gly
145      150      155      160
act gtg aga act ctg agg aca aag cag cgg ata caa cct caa aag acg      647
Thr Val Arg Thr Leu Arg Thr Lys Gln Arg Ile Gln Pro Gln Lys Thr
165      170      175
tct gtc tac att gaa ttg gga tct gat tct tct gaa gat acc gtt aat      695
Ser Val Tyr Ile Glu Leu Gly Ser Asp Ser Ser Glu Asp Thr Val Asn
180      185      190
aag gca act tat tgc agt gtg gga gat caa gaa ttg tta caa atc acc      743
Lys Ala Thr Tyr Cys Ser Val Gly Asp Gln Glu Leu Leu Gln Ile Thr
195      200      205
cct caa gga acc agg gat gaa atc agt ttg gat tct gca aaa aag gct      791
Pro Gln Gly Thr Arg Asp Glu Ile Ser Leu Asp Ser Ala Lys Lys Ala
210      215      220
gct tgt gaa ttt tct gag acg gat gta aca aat act gaa cat cat caa      839
Ala Cys Glu Phe Ser Glu Thr Asp Val Thr Asn Thr Glu His His Gln
225      230      235      240
ccc agt aat aat gat ttg aac acc act gag aag cgt gca gct gag agg      887
Pro Ser Asn Asn Asp Leu Asn Thr Thr Glu Lys Arg Ala Ala Glu Arg
245      250      255

```

A000613

US 6,951,721 B2

133

134

-continued

cat cca gaa aag tat cag ggt agt tct gtt tca aac ttg cat gtg gag His Pro Glu Lys Tyr Gln Gly Ser Ser Val Ser Asn Leu His Val Glu 260 265 270	935
cca tgt ggc aca aat act cat gcc agc tca tta cag cat gag aac agc Pro Cys Gly Thr Asn Thr His Ala Ser Ser Leu Gln His Glu Asn Ser 275 280 285	983
agt tta tta ctc act aaa gac aga atg aat gta gaa aag gct gaa ttc Ser Leu Leu Leu Thr Lys Asp Arg Met Asn Val Glu Lys Ala Glu Phe 290 295 300	1031
tgt aat aaa agc aaa cag cct ggc tta gca agg agc caa cat aac aga Cys Asn Lys Ser Lys Gln Pro Gly Leu Ala Arg Ser Gln His Asn Arg 305 310 315 320	1079
tgg gct gga agt aag gaa aca tgt aat gat agg cgg act ccc agc aca Trp Ala Gly Ser Lys Glu Thr Cys Asn Asp Arg Arg Thr Pro Ser Thr 325 330 335	1127
gaa aaa aag gta gat ctg aat gct gat ccc ctg tgt gag aga aaa gaa Glu Lys Lys Val Asp Leu Asn Ala Asp Pro Leu Cys Glu Arg Lys Glu 340 345 350	1175
tgg aat aag cag aaa ctg cca tgc tca gag aat cct aga gat act gaa Trp Asn Lys Gln Lys Leu Pro Cys Ser Glu Asn Pro Arg Asp Thr Glu 355 360 365	1223
gat gtt cct tgg ata aca cta aat agc agc att cag aaa gtt aat gag Asp Val Pro Trp Ile Thr Leu Asn Ser Ser Ile Gln Lys Val Asn Glu 370 375 380	1271
tgg ttt tcc aga agt gat gaa ctg tta ggt tct gat gac tca cat gat Trp Phe Ser Arg Ser Asp Glu Leu Leu Gly Ser Asp Asp Ser His Asp 385 390 395 400	1319
ggg gag tct gaa tca aat gcc aaa gta gct gat gta ttg gac gtt cta Gly Glu Ser Glu Ser Asn Ala Lys Val Ala Asp Val Leu Asp Val Leu 405 410 415	1367
aat gag gta gat gaa tat tct ggt tct tca gag aaa ata gac tta ctg Asn Glu Val Asp Glu Tyr Ser Gly Ser Ser Glu Lys Ile Asp Leu Leu 420 425 430	1415
gcc agt gat cct cat gag gct tta ata tgt aaa agt gaa aga gtt cac Ala Ser Asp Pro His Glu Ala Leu Ile Cys Lys Ser Glu Arg Val His 435 440 445	1463
tcc aaa tca gta gag agt aat att gaa gac aaa ata ttt ggg aaa acc Ser Lys Ser Val Glu Ser Asn Ile Glu Asp Lys Ile Phe Gly Lys Thr 450 455 460	1511
tat cgg aag aag gca agc ctc ccc aac tta agc cat gta act gaa aat Tyr Arg Lys Lys Ala Ser Leu Pro Asn Leu Ser His Val Thr Glu Asn 465 470 475 480	1559
cta att ata gga gca ttt gtt act gag cca cag ata ata caa gag cgt Leu Ile Ile Gly Ala Phe Val Thr Glu Pro Gln Ile Ile Gln Glu Arg 485 490 495	1607
ccc ctc aca aat aaa tta aag cgt aaa agg aga cct aca tca ggc ctt Pro Leu Thr Asn Lys Leu Lys Arg Lys Arg Arg Pro Thr Ser Gly Leu 500 505 510	1655
cat cct gag gat ttt atc aag aaa gca gat ttg gca gtt caa aag act His Pro Glu Asp Phe Ile Lys Lys Ala Asp Leu Ala Val Gln Lys Thr 515 520 525	1703
cct gaa atg ata aat cag gga act aac caa acg gag cag aat ggt caa Pro Glu Met Ile Asn Gln Gly Thr Asn Gln Thr Glu Gln Asn Gly Gln 530 535 540	1751
gtg atg aat att act aat agt ggt cat gag aat aaa aca aaa ggt gat Val Met Asn Ile Thr Asn Ser Gly His Glu Asn Lys Thr Lys Gly Asp 545 550 555 560	1799
tct att cag aat gag aaa aat cct aac cca ata gaa tca ctc gaa aaa Ser Ile Gln Asn Glu Lys Asn Pro Asn Pro Ile Glu Ser Leu Glu Lys	1847

A000614

US 6,951,721 B2

135

136

-continued

565	570	575	
gaa tct gct ttc aaa acg aaa gct Glu Ser Ala Phe Lys Thr Lys Ala 580	gaa cct ata agc agc agt ata agc Glu Pro Ile Ser Ser Ser Ile Ser 585		1895
aat atg gaa ctc gaa tta aat atc Asn Met Glu Leu Glu Leu Asn Ile 595	cac aat tca aaa gca cct aaa aag Ile His Asn Ser Lys Ala Pro Lys Lys 600		1943
aat agg ctg agg agg aag tct tct Asn Arg Leu Arg Arg Lys Ser Ser 610	acc agg cat att cat gcg ctt gaa Thr Arg His Ile His Ala Leu Glu 615		1991
cta gta gtc agt aga aat cta agc Leu Val Val Ser Arg Asn Leu Ser 625	cca cct aat tgt act gaa ttg caa Pro Pro Asn Cys Thr Glu Leu Gln 630		2039
att gat agt tgt tct agc agt gaa Ile Asp Ser Cys Ser Ser Ser Glu 645	gag ata aag aaa aaa aag tac aac Glu Ile Lys Lys Lys Lys Tyr Asn 650		2087
caa atg cca gtc agg cac agc aga Gln Met Pro Val Arg His Ser Arg 660	aac cta caa ctc atg gaa ggt aaa Asn Leu Gln Leu Met Glu Gly Lys 665		2135
gaa cct gca act gga gcc aag aag Glu Pro Ala Thr Gly Ala Lys Lys 675	agt aac aag cca aat gaa cag aca Ser Asn Lys Pro Asn Glu Gln Thr 680		2183
agt aaa aga cat gac agt gat act Ser Lys Arg His Asp Ser Asp Thr 690	ttc cca gag ctg aag tta aca aat Phe Pro Glu Leu Lys Leu Thr Asn 695		2231
gca cct ggt tct ttt act aag tgt Ala Pro Gly Ser Phe Thr Lys Cys 705	tca aat acc agt gaa ctt aaa gaa Ser Asn Thr Ser Glu Leu Lys Glu 710		2279
ttt gtc aat cct agc ctt cca aga Phe Val Asn Pro Ser Leu Pro Arg 725	gaa gaa aaa gaa gag aaa cta gaa Glu Glu Lys Glu Glu Lys Leu Glu 730		2327
aca gtt aaa gtg tct aat aat gct Thr Val Lys Val Ser Asn Asn Ala 740	gaa gac ccc aaa gat ctc atg tta Glu Asp Pro Lys Asp Leu Met Leu 745		2375
agt gga gaa agg gtt ttg caa act Ser Gly Glu Arg Val Leu Gln Thr 755	gaa aga tct gta gag agt agc agt Glu Arg Ser Val Glu Ser Ser Ser 760		2423
att tca ctg gta cct ggt act gat Ile Ser Leu Val Pro Gly Thr Asp 770	tat ggc act cag gaa agt atc tcg Tyr Gly Thr Gln Glu Ser Ile Ser 775		2471
tta ctg gaa gtt agc act cta ggg Leu Leu Glu Val Ser Thr Leu Gly 785	aag gca aaa aca gaa cca aat aaa Lys Ala Lys Thr Glu Pro Asn Lys 790		2519
tgt gtg agt cag tgt gca gca ttt Cys Val Ser Gln Cys Ala Ala Phe 805	gaa aac ccc aag gga cta att cat Glu Asn Pro Lys Gly Leu Ile His 810		2567
ggt tgt tcc aaa gat aat aga aat Gly Cys Ser Lys Asp Asn Arg Asn 820	gac aca gaa ggc ttt aag tat cca Asp Thr Glu Gly Phe Lys Tyr Pro 825		2615
ttg gga cat gaa gtt aac cac agt Leu Gly His Glu Val Asn His Ser 835	cgg gaa aca agc ata gaa atg gaa Arg Glu Thr Ser Ile Glu Met Glu 840		2663
gaa agt gaa ctt gat gct cag tat Glu Ser Glu Leu Asp Ala Gln Tyr 850	ttg cag aat aca ttc aag gtt tca Leu Gln Asn Thr Phe Lys Val Ser 855		2711
aag cgc cag tca ttt gct ctg ttt Lys Arg Gln Ser Phe Ala Leu Phe 865	tca aat cca gga aat gca gaa gag Ser Asn Pro Gly Asn Ala Glu Glu 870		2759
gaa tgt gca aca ttc tct gcc cac gag tcc tta aag aaa caa agt			2807

A000615

US 6,951,721 B2

137

138

-continued

Glu Cys Ala Thr Phe Ser Ala His Ser Gly Ser Leu Lys Lys Gln Ser	
885 890 895	
cca aaa gtc act ttt gaa tgt gaa caa aag gaa gaa aat caa gga aag	2855
Pro Lys Val Thr Phe Glu Cys Glu Gln Lys Glu Glu Asn Gln Gly Lys	
900 905 910	
aat gag tct aat atc aag cct gta cag aca gtt aat atc act gca ggc	2903
Asn Glu Ser Asn Ile Lys Pro Val Gln Thr Val Asn Ile Thr Ala Gly	
915 920 925	
ttt cct gtg gtt ggt cag aaa gat aag cca gtt gat aat gcc aaa tgt	2951
Phe Pro Val Val Gly Gln Lys Asp Lys Pro Val Asp Asn Ala Lys Cys	
930 935 940	
agt atc aaa gga ggc tct agg ttt tgt cta tca tct cag ttc aga ggc	2999
Ser Ile Lys Gly Gly Ser Arg Phe Cys Leu Ser Ser Gln Phe Arg Gly	
945 950 955 960	
aac gaa act gga ctc att act cca aat aaa cat gga ctt tta caa aac	3047
Asn Glu Thr Gly Leu Ile Thr Pro Asn Lys His Gly Leu Leu Gln Asn	
965 970 975	
cca tat cgt ata cca cca ctt ttt ccc atc aag tca ttt gtt aaa act	3095
Pro Tyr Arg Ile Pro Pro Leu Phe Pro Ile Lys Ser Phe Val Lys Thr	
980 985 990	
aaa tgt aag aaa aat ctg cta gag gaa aac ttt gag gaa cat tca atg	3143
Lys Cys Lys Lys Asn Leu Leu Glu Glu Asn Phe Glu Glu His Ser Met	
995 1000 1005	
tca cct gaa aga gaa atg gga aat gag aac att cca agt aca gtg	3188
Ser Pro Glu Arg Glu Met Gly Asn Glu Asn Ile Pro Ser Thr Val	
1010 1015 1020	
agc aca att agc cgt aat aac att aga gaa aat gtt ttt aaa gga	3233
Ser Thr Ile Ser Arg Asn Asn Ile Arg Glu Asn Val Phe Lys Gly	
1025 1030 1035	
gcc agc tca agc aat att aat gaa gta ggt tcc agt act aat gaa	3278
Ala Ser Ser Ser Asn Ile Asn Glu Val Gly Ser Ser Thr Asn Glu	
1040 1045 1050	
gtg ggc tcc agt att aat gaa ata ggt tcc agt gat gaa aac att	3323
Val Gly Ser Ser Ile Asn Glu Ile Gly Ser Ser Asp Glu Asn Ile	
1055 1060 1065	
caa gca gaa cta ggt aga aac aga ggg cca aaa ttg aat gct atg	3368
Gln Ala Glu Leu Gly Arg Asn Arg Gly Pro Lys Leu Asn Ala Met	
1070 1075 1080	
ctt aga tta ggg gtt ttg caa cct gag gtc tat aaa caa agt ctt	3413
Leu Arg Leu Gly Val Leu Gln Pro Glu Val Tyr Lys Gln Ser Leu	
1085 1090 1095	
cct gga agt aat tgt aag cat cct gaa ata aaa aag caa gaa tat	3458
Pro Gly Ser Asn Cys Lys His Pro Glu Ile Lys Lys Gln Glu Tyr	
1100 1105 1110	
gaa gaa gta gtt cag act gtt aat aca gat ttc tct cca tat ctg	3503
Glu Glu Val Val Gln Thr Val Asn Thr Asp Phe Ser Pro Tyr Leu	
1115 1120 1125	
att tca gat aac tta gaa cag cct atg gga agt agt cat gca tct	3548
Ile Ser Asp Asn Leu Glu Gln Pro Met Gly Ser Ser His Ala Ser	
1130 1135 1140	
cag gtt tgt tct gag aca cct gat gac ctg tta gat gat ggt gaa	3593
Gln Val Cys Ser Glu Thr Pro Asp Asp Leu Leu Asp Asp Gly Glu	
1145 1150 1155	
ata aag gaa gat act agt ttt gct gaa aat gac att aag gaa agt	3638
Ile Lys Glu Asp Thr Ser Phe Ala Glu Asn Asp Ile Lys Glu Ser	
1160 1165 1170	
tct gct gtt ttt agc aaa agc gtc cag aga gga gag ctt agc agg	3683
Ser Ala Val Phe Ser Lys Ser Val Gln Arg Gly Glu Leu Ser Arg	
1175 1180 1185	

A000616

US 6,951,721 B2

139

140

-continued

agt cct	agc cct ttc acc cat	aca cat ttg gct cag	ggg tac cga	3728
Ser Pro	Ser Pro Phe Thr His	Thr His Leu Ala Gln	Gly Tyr Arg	
1190	1195	1200		
aga ggg	gcc aag aaa tta gag	tcc tca gaa gag aac	tta tct agt	3773
Arg Gly	Ala Lys Lys Leu Glu	Ser Ser Glu Glu Asn	Leu Ser Ser	
1205	1210	1215		
gag gat	gaa gag ctt ccc tgc	ttc caa cac ttg tta	ttt ggt aaa	3818
Glu Asp	Glu Glu Leu Pro Cys	Phe Gln His Leu Leu	Phe Gly Lys	
1220	1225	1230		
gta aac	aat ata cct tct cag	tct act agg cat agc	acc gtt gct	3863
Val Asn	Asn Ile Pro Ser Gln	Ser Thr Arg His Ser	Thr Val Ala	
1235	1240	1245		
acc gag	tgt ctg tct aag aac	aca gag gag aat tta	tta tca ttg	3908
Thr Glu	Cys Leu Ser Lys Asn	Thr Glu Glu Asn Leu	Leu Ser Leu	
1250	1255	1260		
aag aat	agc tta aat gac tgc	agt aac cag gta ata	ttg gca aag	3953
Lys Asn	Ser Leu Asn Asp Cys	Ser Asn Gln Val Ile	Leu Ala Lys	
1265	1270	1275		
gca tct	cag gaa cat cac ctt	agt gag gaa aca aaa	tgt tct gct	3998
Ala Ser	Gln Glu His His Leu	Ser Glu Glu Thr Lys	Cys Ser Ala	
1280	1285	1290		
agc ttg	ttt tct tca cag tgc	agt gaa ttg gaa gac	ttg act gca	4043
Ser Leu	Phe Ser Ser Gln Cys	Ser Glu Leu Glu Asp	Leu Thr Ala	
1295	1300	1305		
aat aca	aac acc cag gat cct	ttc ttg att ggt tct	tcc aaa caa	4088
Asn Thr	Asn Thr Gln Asp Pro	Phe Leu Ile Gly Ser	Ser Lys Gln	
1310	1315	1320		
atg agg	cat cag tct gaa agc	cag gga gtt ggt ctg	agt gac aag	4133
Met Arg	His Gln Ser Glu Ser	Gln Gly Val Gly Leu	Ser Asp Lys	
1325	1330	1335		
gaa ttg	gtt tca gat gat gaa	gaa aga gga acg ggc	ttg gaa gaa	4178
Glu Leu	Val Ser Asp Asp Glu	Glu Arg Gly Thr Gly	Leu Glu Glu	
1340	1345	1350		
aat aat	caa gaa gag caa agc	atg gat tca aac tta	ggg gaa gca	4223
Asn Asn	Gln Glu Glu Gln Ser	Met Asp Ser Asn Leu	Gly Glu Ala	
1355	1360	1365		
gca tct	ggg tgt gag agt gaa	aca agc gtc tct gaa	gac tgc tca	4268
Ala Ser	Gly Cys Glu Ser Glu	Thr Ser Val Ser Glu	Asp Cys Ser	
1370	1375	1380		
ggg cta	tcc tct cag agt gac	att tta acc act cag	cag agg gat	4313
Gly Leu	Ser Ser Gln Ser Asp	Ile Leu Thr Thr Gln	Gln Arg Asp	
1385	1390	1395		
acc atg	caa cat aac ctg ata	aag ctc cag cag gaa	atg gct gaa	4358
Thr Met	Gln His Asn Leu Ile	Lys Leu Gln Gln Glu	Met Ala Glu	
1400	1405	1410		
cta gaa	gct gtg tta gaa cag	cat ggg agc cag cct	tct aac agc	4403
Leu Glu	Ala Val Leu Glu Gln	His Gly Ser Gln Pro	Ser Asn Ser	
1415	1420	1425		
tac cct	tcc atc ata agt gac	tct tct gcc ctt gag	gac ctg cga	4448
Tyr Pro	Ser Ile Ile Ser Asp	Ser Ser Ala Leu Glu	Asp Leu Arg	
1430	1435	1440		
aat cca	gaa caa agc aca tca	gaa aaa gca gta tta	act tca cag	4493
Asn Pro	Glu Gln Ser Thr Ser	Glu Lys Ala Val Leu	Thr Ser Gln	
1445	1450	1455		
aaa agt	agt gaa tac cct ata	agc cag aat cca gaa	ggc ctt tct	4538
Lys Ser	Ser Glu Tyr Pro Ile	Ser Gln Asn Pro Glu	Gly Leu Ser	
1460	1465	1470		
gct gac	aag ttt gag gtg tct	gca gat agt tct acc	agt aaa aat	4583
Ala Asp	Lys Phe Glu Val Ser	Ala Asp Ser Ser Thr	Ser Lys Asn	
1475	1480	1485		

A000617

US 6,951,721 B2

141

142

-continued

aaa gaa cca gga gtg gaa agg tca tcc cct tct aaa tgc cca tca 4628 Lys Glu Pro Gly Val Glu Arg Ser Ser Pro Ser Lys Cys Pro Ser 1490 1495 1500
tta gat gat agg tgg tac atg cac agt tgc tct ggg agt ctt cag 4673 Leu Asp Asp Arg Trp Tyr Met His Ser Cys Ser Gly Ser Leu Gln 1505 1510 1515
aat aga aac tac cca tct caa gag gag ctc att aag gtt gtt gat 4718 Asn Arg Asn Tyr Pro Ser Glu Glu Leu Ile Lys Val Val Asp 1520 1525 1530
gtg gag gag caa cag ctg gaa gag tct ggg cca cac gat ttg acg 4763 Val Glu Glu Gln Gln Leu Glu Glu Ser Gly Pro His Asp Leu Thr 1535 1540 1545
gaa aca tct tac ttg cca agg caa gat cta gag gga acc cct tac 4808 Glu Thr Ser Tyr Leu Pro Arg Gln Asp Leu Glu Gly Thr Pro Tyr 1550 1555 1560
ctg gaa tct gga atc agc ctc ttc tct gat gac cct gaa tct gat 4853 Leu Glu Ser Gly Ile Ser Leu Phe Ser Asp Asp Pro Glu Ser Asp 1565 1570 1575
cct tct gaa gac aga gcc cca gag tca gct cgt gtt ggc aac ata 4898 Pro Ser Glu Asp Arg Ala Pro Glu Ser Ala Arg Val Gly Asn Ile 1580 1585 1590
cca tct tca acc tct gca ttg aaa gtt ccc caa ttg aaa gtt gca 4943 Pro Ser Ser Thr Ser Ala Leu Lys Val Pro Gln Leu Lys Val Ala 1595 1600 1605
gaa tct gcc cag ggt cca gct gct gct cat act act gat act gct 4988 Glu Ser Ala Gln Gly Pro Ala Ala Ala His Thr Thr Asp Thr Ala 1610 1615 1620
ggg tat aat gca atg gaa gaa agt gtg agc agg gag aag cca gaa 5033 Gly Tyr Asn Ala Met Glu Glu Ser Val Ser Arg Glu Lys Pro Glu 1625 1630 1635
ttg aca gct tca aca gaa agg gtc aac aaa aga atg tcc atg gtg 5078 Leu Thr Ala Ser Thr Glu Arg Val Asn Lys Arg Met Ser Met Val 1640 1645 1650
gtg tct ggc ctg acc cca gaa gaa ttt atg ctc gtg tac aag ttt 5123 Val Ser Gly Leu Thr Pro Glu Glu Phe Met Leu Val Tyr Lys Phe 1655 1660 1665
gcc aga aaa cac cac atc act tta act aat cta att act gaa gag 5168 Ala Arg Lys His His Ile Thr Leu Thr Asn Leu Ile Thr Glu Glu 1670 1675 1680
act act cat gtt gtt atg aaa aca gat gct gag ttt gtg tgt gaa 5213 Thr Thr His Val Val Met Lys Thr Asp Ala Glu Phe Val Cys Glu 1685 1690 1695
cgg aca ctg aaa tat ttt cta gga att gcg gga gga aaa tgg gta 5258 Arg Thr Leu Lys Tyr Phe Leu Gly Ile Ala Gly Gly Lys Trp Val 1700 1705 1710
gtt agc tat ttc tgg gtg acc cag tct att aaa gaa aga aaa atg 5303 Val Ser Tyr Phe Trp Val Thr Gln Ser Ile Lys Glu Arg Lys Met 1715 1720 1725
ctg aat gag cat gat ttt gaa gtc aga gga gat gtg gtc aat gga 5348 Leu Asn Glu His Asp Phe Glu Val Arg Gly Asp Val Val Asn Gly 1730 1735 1740
aga aac cac caa ggt cca aag cga gca aga gaa tcc cag gac aga 5393 Arg Asn His Gln Gly Pro Lys Arg Ala Arg Glu Ser Gln Asp Arg 1745 1750 1755
aag atc ttc agg ggg cta gaa atc tgt tgc tat ggg ccc ttc acc 5438 Lys Ile Phe Arg Gly Leu Glu Ile Cys Cys Tyr Gly Pro Phe Thr 1760 1765 1770
aac atg ccc aca gat caa ctg gaa tgg atg gta cag ctg tgt ggt 5483 Asn Met Pro Thr Asp Gln Leu Glu Trp Met Val Gln Leu Cys Gly

A000618

US 6,951,721 B2

143

144

-continued

1775	1780	1785	
gct tct gtg gtg aag gag ctt tca tca ttc acc ctt ggc aca ggt			5528
Ala Ser Val Val Lys Glu Leu Ser Ser Phe Thr Leu Gly Thr Gly			
1790	1795	1800	
gtc cac cca att gtg gtt gtg cag cca gat gcc tgg aca gag gac			5573
Val His Pro Ile Val Val Val Gln Pro Asp Ala Trp Thr Glu Asp			
1805	1810	1815	
aat ggc ttc cat gca att ggg cag atg tgt gag gca cct gtg gtg			5618
Asn Gly Phe His Ala Ile Gly Gln Met Cys Glu Ala Pro Val Val			
1820	1825	1830	
acc cga gag tgg gtg ttg gac agt gta gca ctc tac cag tgc cag			5663
Thr Arg Glu Trp Val Leu Asp Ser Val Ala Leu Tyr Gln Cys Gln			
1835	1840	1845	
gag ctg gac acc tac ctg ata ccc cag atc ccc cac agc cac tac			5708
Glu Leu Asp Thr Tyr Leu Ile Pro Gln Ile Pro His Ser His Tyr			
1850	1855	1860	
tga			5711

<210> SEQ ID NO 266
 <211> LENGTH: 1863
 <212> TYPE: PRT
 <213> ORGANISM: Homo sapiens

<400> SEQUENCE: 266

Met Asp Leu Ser Ala Leu Arg Val Glu Glu Val Gln Asn Val Ile Asn	
1	15
Ala Met Gln Lys Ile Leu Glu Cys Pro Ile Cys Leu Glu Leu Ile Lys	
20	30
Glu Pro Val Ser Thr Lys Cys Asp His Ile Phe Cys Lys Phe Cys Met	
35	45
Leu Lys Leu Leu Asn Gln Lys Lys Gly Pro Ser Gln Cys Pro Leu Cys	
50	60
Lys Asn Asp Ile Thr Lys Arg Ser Leu Gln Glu Ser Thr Arg Phe Ser	
65	80
Gln Leu Val Glu Glu Leu Leu Lys Ile Ile Cys Ala Phe Gln Leu Asp	
85	95
Thr Gly Leu Glu Tyr Ala Asn Ser Tyr Asn Phe Ala Lys Lys Glu Asn	
100	110
Asn Ser Pro Glu His Leu Lys Asp Glu Val Ser Ile Ile Gln Ser Met	
115	125
Gly Tyr Arg Asn Arg Ala Lys Arg Leu Leu Gln Ser Glu Pro Glu Asn	
130	140
Pro Ser Leu Gln Glu Thr Ser Leu Ser Val Gln Leu Ser Asn Leu Gly	
145	160
Thr Val Arg Thr Leu Arg Thr Lys Gln Arg Ile Gln Pro Gln Lys Thr	
165	175
Ser Val Tyr Ile Glu Leu Gly Ser Asp Ser Ser Glu Asp Thr Val Asn	
180	190
Lys Ala Thr Tyr Cys Ser Val Gly Asp Gln Glu Leu Leu Gln Ile Thr	
195	205
Pro Gln Gly Thr Arg Asp Glu Ile Ser Leu Asp Ser Ala Lys Lys Ala	
210	220
Ala Cys Glu Phe Ser Glu Thr Asp Val Thr Asn Thr Glu His His Gln	
225	240
Pro Ser Asn Asn Asp Leu Asn Thr Thr Glu Lys Arg Ala Ala Glu Arg	
245	255

A000619

US 6,951,721 B2

145

146

-continued

His Pro Glu Lys Tyr Gln Gly Ser Ser Val Ser Asn Leu His Val Glu
 260 265 270
 Pro Cys Gly Thr Asn Thr His Ala Ser Ser Leu Gln His Glu Asn Ser
 275 280 285
 Ser Leu Leu Leu Thr Lys Asp Arg Met Asn Val Glu Lys Ala Glu Phe
 290 295 300
 Cys Asn Lys Ser Lys Gln Pro Gly Leu Ala Arg Ser Gln His Asn Arg
 305 310 315 320
 Trp Ala Gly Ser Lys Glu Thr Cys Asn Asp Arg Arg Thr Pro Ser Thr
 325 330 335
 Glu Lys Lys Val Asp Leu Asn Ala Asp Pro Leu Cys Glu Arg Lys Glu
 340 345 350
 Trp Asn Lys Gln Lys Leu Pro Cys Ser Glu Asn Pro Arg Asp Thr Glu
 355 360 365
 Asp Val Pro Trp Ile Thr Leu Asn Ser Ser Ile Gln Lys Val Asn Glu
 370 375 380
 Trp Phe Ser Arg Ser Asp Glu Leu Leu Gly Ser Asp Asp Ser His Asp
 385 390 395 400
 Gly Glu Ser Glu Ser Asn Ala Lys Val Ala Asp Val Leu Asp Val Leu
 405 410 415
 Asn Glu Val Asp Glu Tyr Ser Gly Ser Ser Glu Lys Ile Asp Leu Leu
 420 425 430
 Ala Ser Asp Pro His Glu Ala Leu Ile Cys Lys Ser Glu Arg Val His
 435 440 445
 Ser Lys Ser Val Glu Ser Asn Ile Glu Asp Lys Ile Phe Gly Lys Thr
 450 455 460
 Tyr Arg Lys Lys Ala Ser Leu Pro Asn Leu Ser His Val Thr Glu Asn
 465 470 475 480
 Leu Ile Ile Gly Ala Phe Val Thr Glu Pro Gln Ile Ile Gln Glu Arg
 485 490 495
 Pro Leu Thr Asn Lys Leu Lys Arg Lys Arg Arg Pro Thr Ser Gly Leu
 500 505 510
 His Pro Glu Asp Phe Ile Lys Lys Ala Asp Leu Ala Val Gln Lys Thr
 515 520 525
 Pro Glu Met Ile Asn Gln Gly Thr Asn Gln Thr Glu Gln Asn Gly Gln
 530 535 540
 Val Met Asn Ile Thr Asn Ser Gly His Glu Asn Lys Thr Lys Gly Asp
 545 550 555 560
 Ser Ile Gln Asn Glu Lys Asn Pro Asn Pro Ile Glu Ser Leu Glu Lys
 565 570 575
 Glu Ser Ala Phe Lys Thr Lys Ala Glu Pro Ile Ser Ser Ser Ile Ser
 580 585 590
 Asn Met Glu Leu Glu Leu Asn Ile His Asn Ser Lys Ala Pro Lys Lys
 595 600 605
 Asn Arg Leu Arg Arg Lys Ser Ser Thr Arg His Ile His Ala Leu Glu
 610 615 620
 Leu Val Val Ser Arg Asn Leu Ser Pro Pro Asn Cys Thr Glu Leu Gln
 625 630 635 640
 Ile Asp Ser Cys Ser Ser Ser Glu Glu Ile Lys Lys Lys Lys Tyr Asn
 645 650 655
 Gln Met Pro Val Arg His Ser Arg Asn Leu Gln Leu Met Glu Gly Lys
 660 665 670

A000620

147

-continued

A000621

US 6,951,721 B2

149

150

-continued

1085	1090	1095
Pro Gly Ser Asn Cys Lys His	Pro Glu Ile Lys Lys	Gln Glu Tyr
1100	1105	1110
Glu Glu Val Val Gln Thr Val	Asn Thr Asp Phe Ser	Pro Tyr Leu
1115	1120	1125
Ile Ser Asp Asn Leu Glu Gln	Pro Met Gly Ser Ser	His Ala Ser
1130	1135	1140
Gln Val Cys Ser Glu Thr Pro	Asp Asp Leu Leu Asp	Asp Gly Glu
1145	1150	1155
Ile Lys Glu Asp Thr Ser Phe	Ala Glu Asn Asp Ile	Lys Glu Ser
1160	1165	1170
Ser Ala Val Phe Ser Lys Ser	Val Gln Arg Gly Glu	Leu Ser Arg
1175	1180	1185
Ser Pro Ser Pro Phe Thr His	Thr His Leu Ala Gln	Gly Tyr Arg
1190	1195	1200
Arg Gly Ala Lys Lys Leu Glu	Ser Ser Glu Glu Asn	Leu Ser Ser
1205	1210	1215
Glu Asp Glu Glu Leu Pro Cys	Phe Gln His Leu Leu	Phe Gly Lys
1220	1225	1230
Val Asn Asn Ile Pro Ser Gln	Ser Thr Arg His Ser	Thr Val Ala
1235	1240	1245
Thr Glu Cys Leu Ser Lys Asn	Thr Glu Glu Asn Leu	Leu Ser Leu
1250	1255	1260
Lys Asn Ser Leu Asn Asp Cys	Ser Asn Gln Val Ile	Leu Ala Lys
1265	1270	1275
Ala Ser Gln Glu His His Leu	Ser Glu Glu Thr Lys	Cys Ser Ala
1280	1285	1290
Ser Leu Phe Ser Ser Gln Cys	Ser Glu Leu Glu Asp	Leu Thr Ala
1295	1300	1305
Asn Thr Asn Thr Gln Asp Pro	Phe Leu Ile Gly Ser	Ser Lys Gln
1310	1315	1320
Met Arg His Gln Ser Glu Ser	Gln Gly Val Gly Leu	Ser Asp Lys
1325	1330	1335
Glu Leu Val Ser Asp Asp Glu	Glu Arg Gly Thr Gly	Leu Glu Glu
1340	1345	1350
Asn Asn Gln Glu Glu Gln Ser	Met Asp Ser Asn Leu	Gly Glu Ala
1355	1360	1365
Ala Ser Gly Cys Glu Ser Glu	Thr Ser Val Ser Glu	Asp Cys Ser
1370	1375	1380
Gly Leu Ser Ser Gln Ser Asp	Ile Leu Thr Thr Gln	Gln Arg Asp
1385	1390	1395
Thr Met Gln His Asn Leu Ile	Lys Leu Gln Gln Glu	Met Ala Glu
1400	1405	1410
Leu Glu Ala Val Leu Glu Gln	His Gly Ser Gln Pro	Ser Asn Ser
1415	1420	1425
Tyr Pro Ser Ile Ile Ser Asp	Ser Ser Ala Leu Glu	Asp Leu Arg
1430	1435	1440
Asn Pro Glu Gln Ser Thr Ser	Glu Lys Ala Val Leu	Thr Ser Gln
1445	1450	1455
Lys Ser Ser Glu Tyr Pro Ile	Ser Gln Asn Pro Glu	Gly Leu Ser
1460	1465	1470
Ala Asp Lys Phe Glu Val Ser	Ala Asp Ser Ser Thr	Ser Lys Asn
1475	1480	1485

A000622

US 6,951,721 B2

151

152

-continued

Lys	Glu	Pro	Gly	Val	Glu	Arg	Ser	Ser	Pro	Ser	Lys	Cys	Pro	Ser
1490						1495					1500			
Leu	Asp	Asp	Arg	Trp	Tyr	Met	His	Ser	Cys	Ser	Gly	Ser	Leu	Gln
1505						1510					1515			
Asn	Arg	Asn	Tyr	Pro	Ser	Gln	Glu	Glu	Leu	Ile	Lys	Val	Val	Asp
1520						1525					1530			
Val	Glu	Glu	Gln	Gln	Leu	Glu	Glu	Ser	Gly	Pro	His	Asp	Leu	Thr
1535						1540					1545			
Glu	Thr	Ser	Tyr	Leu	Pro	Arg	Gln	Asp	Leu	Glu	Gly	Thr	Pro	Tyr
1550						1555					1560			
Leu	Glu	Ser	Gly	Ile	Ser	Leu	Phe	Ser	Asp	Asp	Pro	Glu	Ser	Asp
1565						1570					1575			
Pro	Ser	Glu	Asp	Arg	Ala	Pro	Glu	Ser	Ala	Arg	Val	Gly	Asn	Ile
1580						1585					1590			
Pro	Ser	Ser	Thr	Ser	Ala	Leu	Lys	Val	Pro	Gln	Leu	Lys	Val	Ala
1595						1600					1605			
Glu	Ser	Ala	Gln	Gly	Pro	Ala	Ala	Ala	His	Thr	Thr	Asp	Thr	Ala
1610						1615					1620			
Gly	Tyr	Asn	Ala	Met	Glu	Glu	Ser	Val	Ser	Arg	Glu	Lys	Pro	Glu
1625						1630					1635			
Leu	Thr	Ala	Ser	Thr	Glu	Arg	Val	Asn	Lys	Arg	Met	Ser	Met	Val
1640						1645					1650			
Val	Ser	Gly	Leu	Thr	Pro	Glu	Glu	Phe	Met	Leu	Val	Tyr	Lys	Phe
1655						1660					1665			
Ala	Arg	Lys	His	His	Ile	Thr	Leu	Thr	Asn	Leu	Ile	Thr	Glu	Glu
1670						1675					1680			
Thr	Thr	His	Val	Val	Met	Lys	Thr	Asp	Ala	Glu	Phe	Val	Cys	Glu
1685						1690					1695			
Arg	Thr	Leu	Lys	Tyr	Phe	Leu	Gly	Ile	Ala	Gly	Gly	Lys	Trp	Val
1700						1705					1710			
Val	Ser	Tyr	Phe	Trp	Val	Thr	Gln	Ser	Ile	Lys	Glu	Arg	Lys	Met
1715						1720					1725			
Leu	Asn	Glu	His	Asp	Phe	Glu	Val	Arg	Gly	Asp	Val	Val	Asn	Gly
1730						1735					1740			
Arg	Asn	His	Gln	Gly	Pro	Lys	Arg	Ala	Arg	Glu	Ser	Gln	Asp	Arg
1745						1750					1755			
Lys	Ile	Phe	Arg	Gly	Leu	Glu	Ile	Cys	Cys	Tyr	Gly	Pro	Phe	Thr
1760						1765					1770			
Asn	Met	Pro	Thr	Asp	Gln	Leu	Glu	Trp	Met	Val	Gln	Leu	Cys	Gly
1775						1780					1785			
Ala	Ser	Val	Val	Lys	Glu	Leu	Ser	Ser	Phe	Thr	Leu	Gly	Thr	Gly
1790						1795					1800			
Val	His	Pro	Ile	Val	Val	Val	Gln	Pro	Asp	Ala	Trp	Thr	Glu	Asp
1805						1810					1815			
Asn	Gly	Phe	His	Ala	Ile	Gly	Gln	Met	Cys	Glu	Ala	Pro	Val	Val
1820						1825					1830			
Thr	Arg	Glu	Trp	Val	Leu	Asp	Ser	Val	Ala	Leu	Tyr	Gln	Cys	Gln
1835						1840					1845			
Glu	Leu	Asp	Thr	Tyr	Leu	Ile	Pro	Gln	Ile	Pro	His	Ser	His	Tyr
1850						1855					1860			

<210> SEQ ID NO 267

<211> LENGTH: 5711

A000623

US 6,951,721 B2

153

154

-continued

```

<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (120)..(5711)
<223> OTHER INFORMATION: omi3 sequences

<400> SEQUENCE: 267

agctcgctga gacttctctgg accccgcacc aggctgtggg gtttctcaga taactggggc 60
cctgcgcctca ggaggccttc accctctgct ctgggtaaag ttcattggaa cagaaagaa 119
atg gat tta tct gct ctt cgc gtt gaa gaa gta caa aat gtc att aat 167
Met Asp Leu Ser Ala Leu Arg Val Glu Glu Val Gln Asn Val Ile Asn
1 5 10 15
gct atg cag aaa atc tta gag tgt ccc atc tgt ctg gag ttg atc aag 215
Ala Met Gln Lys Ile Leu Glu Cys Pro Ile Cys Leu Glu Leu Ile Lys
20 25 30
gaa cct gtc tcc aca aag tgt gac cac ata ttt tgc aaa ttt tgc atg 263
Glu Pro Val Ser Thr Lys Cys Asp His Ile Phe Cys Lys Phe Cys Met
35 40 45
ctg aaa ctt ctg aac cag aag aaa ggg cct tca cag tgt cct tta tgt 311
Leu Lys Leu Leu Asn Gln Lys Lys Gly Pro Ser Gln Cys Pro Leu Cys
50 55 60
aag aat gat ata acc aaa agg agc cta caa gaa agt acg aga ttt agt 359
Lys Asn Asp Ile Thr Lys Arg Ser Leu Gln Glu Ser Thr Arg Phe Ser
65 70 75 80
caa ctt gtt gaa gag cta ttg aaa atc att tgt gct ttt cag ctt gac 407
Gln Leu Val Glu Glu Leu Leu Lys Ile Ile Cys Ala Phe Gln Leu Asp
85 90 95
aca ggt ttg gag tat gca aac agc tat aat ttt gca aaa aag gaa aat 455
Thr Gly Leu Glu Tyr Ala Asn Ser Tyr Asn Phe Ala Lys Lys Glu Asn
100 105 110
aac tct cct gaa cat cta aaa gat gaa gtt tct atc atc caa agt atg 503
Asn Ser Pro Glu His Leu Lys Asp Glu Val Ser Ile Ile Gln Ser Met
115 120 125
ggc tac aga aac cgt gcc aaa aga ctt cta cag agt gaa ccc gaa aat 551
Gly Tyr Arg Asn Arg Ala Lys Arg Leu Leu Gln Ser Glu Pro Glu Asn
130 135 140
cct tcc ttg cag gaa acc agt ctg agt gtc caa ctg tct aac ctt gga 599
Pro Ser Leu Gln Glu Thr Ser Leu Ser Val Gln Leu Ser Asn Leu Gly
145 150 155 160
act gtg aga act ctg agg aca aag cag cgg ata caa cct caa aag acg 647
Thr Val Arg Thr Leu Arg Thr Lys Gln Arg Ile Gln Pro Gln Lys Thr
165 170 175
tct gtc tac att gaa ttg gga tct gat tct tct gaa gat acc gtt aat 695
Ser Val Tyr Ile Glu Leu Gly Ser Asp Ser Ser Glu Asp Thr Val Asn
180 185 190
aag gca act tat tgc agt gtg gga gat caa gaa ttg tta caa atc acc 743
Lys Ala Thr Tyr Cys Ser Val Gly Asp Gln Glu Leu Leu Gln Ile Thr
195 200 205
cct caa gga acc agg gat gaa atc agt ttg gat tct gca aaa aag gct 791
Pro Gln Gly Thr Arg Asp Glu Ile Ser Leu Asp Ser Ala Lys Lys Ala
210 215 220
gct tgt gaa ttt tct gag acg gat gta aca aat act gaa cat cat caa 839
Ala Cys Glu Phe Ser Glu Thr Asp Val Thr Asn Thr Glu His His Gln
225 230 235 240
ccc agt aat aat gat ttg aac acc act gag aag cgt gca gct gag agg 887
Pro Ser Asn Asn Asp Leu Asn Thr Thr Glu Lys Arg Ala Ala Glu Arg
245 250 255
cat cca gaa aag tat cag ggt agt tct gtt tca aac ttg cat gtg gag 935
His Pro Glu Lys Tyr Gln Gly Ser Ser Val Ser Asn Leu His Val Glu

```

A000624

US 6,951,721 B2

155

156

-continued

260	265	270	
cca tgt ggc aca aat act cat gcc agc tca tta cag cat gag aac agc Pro Cys Gly Thr Asn Thr His Ala Ser Ser Leu Gln His Glu Asn Ser 275 280 285			983
agt tta tta ctc act aaa gac aga atg aat gta gaa aag gct gaa ttc Ser Leu Leu Leu Thr Lys Asp Arg Met Asn Val Glu Lys Ala Glu Phe 290 295 300			1031
tgt aat aaa agc aaa cag cct gcc tta gca agg agc caa cat aac aga Cys Asn Lys Ser Lys Gln Pro Gly Leu Ala Arg Ser Gln His Asn Arg 305 310 315 320			1079
tgg gct gga agt aag gaa aca tgt aat gat agg cgg act ccc agc aca Trp Ala Gly Ser Lys Glu Thr Cys Asn Asp Arg Arg Thr Pro Ser Thr 325 330 335			1127
gaa aaa aag gta gat ctg aat gct gat ccc ctg tgt gag aga aaa gaa Glu Lys Lys Val Asp Leu Asn Ala Asp Pro Leu Cys Glu Arg Lys Glu 340 345 350			1175
tgg aat aag cag aaa ctg cca tgc tca gag aat cct aga gat act gaa Trp Asn Lys Gln Lys Leu Pro Cys Ser Glu Asn Pro Arg Asp Thr Glu 355 360 365			1223
gat gtt cct tgg ata aca cta aat agc agc att cag aaa gtt aat gag Asp Val Pro Trp Ile Thr Leu Asn Ser Ser Ile Gln Lys Val Asn Glu 370 375 380			1271
tgg ttt tcc aga agt gat gaa ctg tta ggt tct gat gac tca cat gat Trp Phe Ser Arg Ser Asp Glu Leu Leu Gly Ser Asp Asp Ser His Asp 385 390 395 400			1319
ggg gag tct gaa tca aat gcc aaa gta gct gat gta ttg gac gtt cta Gly Glu Ser Glu Ser Asn Ala Lys Val Ala Asp Val Leu Asp Val Leu 405 410 415			1367
aat gag gta gat gaa tat tct ggt tct tca gag aaa ata gac tta ctg Asn Glu Val Asp Glu Tyr Ser Gly Ser Ser Glu Lys Ile Asp Leu Leu 420 425 430			1415
gcc agt gat cct cat gag gct tta ata tgt aaa agt gaa aga gtt cac Ala Ser Asp Pro His Glu Ala Leu Ile Cys Lys Ser Glu Arg Val His 435 440 445			1463
tcc aaa tca gta gag agt aat att gaa gac aaa ata ttt ggg aaa acc Ser Lys Ser Val Glu Ser Asn Ile Glu Asp Lys Ile Phe Gly Lys Thr 450 455 460			1511
tat cgg aag aag gca agc ctc ccc aac tta agc cat gta act gaa aat Tyr Arg Lys Lys Ala Ser Leu Pro Asn Leu Ser His Val Thr Glu Asn 465 470 475 480			1559
cta att ata gga gca ttt gtt act gag cca cag ata ata caa gag cgt Leu Ile Ile Gly Ala Phe Val Thr Glu Pro Gln Ile Ile Gln Glu Arg 485 490 495			1607
ccc ctc aca aat aaa tta aag cgt aaa agg aga cct aca tca ggc ctt Pro Leu Thr Asn Lys Leu Lys Arg Lys Arg Arg Pro Thr Ser Gly Leu 500 505 510			1655
cat cct gag gat ttt atc aag aaa gca gat ttg gca gtt caa aag act His Pro Glu Asp Phe Ile Lys Lys Ala Asp Leu Ala Val Gln Lys Thr 515 520 525			1703
cct gaa atg ata aat cag gga act aac caa acg gag cag aat ggt caa Pro Glu Met Ile Asn Gln Gly Thr Asn Gln Thr Glu Gln Asn Gly Gln 530 535 540			1751
gtg atg aat att act aat agt ggt cat gag aat aaa aca aaa ggt gat Val Met Asn Ile Thr Asn Ser Gly His Glu Asn Lys Thr Lys Gly Asp 545 550 555 560			1799
tct att cag aat gag aaa aat cct aac cca ata gaa tca ctc gaa aaa Ser Ile Gln Asn Glu Lys Asn Pro Asn Pro Ile Glu Ser Leu Glu Lys 565 570 575			1847
gaa tct gct ttc aaa acg aaa gct gaa cct ata agc agc agt ata agc			1895

A000625

US 6,951,721 B2

157

158

-continued

Glu Ser Ala Phe Lys Thr Lys Ala Glu Pro Ile Ser Ser Ser Ile Ser	
580 585 590	
aat atg gaa ctc gaa tta aat atc cac aat tca aaa gca cct aaa aag	1943
Asn Met Glu Leu Glu Leu Asn Ile His Asn Ser Lys Ala Pro Lys Lys	
595 600 605	
aat agg ctg agg agg aag tct tct acc agg cat att cat gcg ctt gaa	1991
Asn Arg Leu Arg Arg Lys Ser Ser Thr Arg His Ile His Ala Leu Glu	
610 615 620	
cta gta gtc agt aga aat cta agc cca cct aat tgt act gaa ttg caa	2039
Leu Val Val Ser Arg Asn Leu Ser Pro Pro Asn Cys Thr Glu Leu Gln	
625 630 635	
att gat agt tgt tct agc agt gaa gag ata aag aaa aag tac aac	2087
Ile Asp Ser Cys Ser Ser Ser Glu Glu Ile Lys Lys Lys Tyr Asn	
645 650 655	
caa atg cca gtc agg cac agc aga aac cta caa ctc atg gaa ggt aaa	2135
Gln Met Pro Val Arg His Ser Arg Asn Leu Gln Leu Met Glu Gly Lys	
660 665 670	
gaa cct gca act gga gcc aag aag agt aac aag cca aat gaa cag aca	2183
Glu Pro Ala Thr Gly Ala Lys Lys Ser Asn Lys Pro Asn Glu Gln Thr	
675 680 685	
agt aaa aga cat gac agc gat act ttc cca gag ctg aag tta aca aat	2231
Ser Lys Arg His Asp Ser Asp Thr Phe Pro Glu Leu Lys Leu Thr Asn	
690 695 700	
gca cct ggt tct ttt act aag tgt tca aat acc agt gaa ctt aaa gaa	2279
Ala Pro Gly Ser Phe Thr Lys Cys Ser Asn Thr Ser Glu Leu Lys Glu	
705 710 715 720	
ttt gtc aat cct agc ctt cca aga gaa gaa aaa gaa gag aaa cta gaa	2327
Phe Val Asn Pro Ser Leu Pro Arg Glu Glu Lys Glu Glu Lys Leu Glu	
725 730 735	
aca gtt aaa gtg tct aat aat gct gaa gac ccc aaa gat ctc atg tta	2375
Thr Val Lys Val Ser Asn Asn Ala Glu Asp Pro Lys Asp Leu Met Leu	
740 745 750	
agt gga gaa agg gtt ttg caa act gaa aga tct gta gag agt agc agt	2423
Ser Gly Glu Arg Val Leu Gln Thr Glu Arg Ser Val Glu Ser Ser Ser	
755 760 765	
att tca ttg gta cct ggt act gat tat ggc act cag gaa agt atc tcg	2471
Ile Ser Leu Val Pro Gly Thr Asp Tyr Gly Thr Gln Glu Ser Ile Ser	
770 775 780	
tta ctg gaa gtt agc act cta ggg aag gca aaa aca gaa cca aat aaa	2519
Leu Leu Glu Val Ser Thr Leu Gly Lys Ala Lys Thr Glu Pro Asn Lys	
785 790 795 800	
tgt gtg agt cag tgt gca gca ttt gaa aac ccc aag gga cta att cat	2567
Cys Val Ser Gln Cys Ala Ala Phe Glu Asn Pro Lys Gly Leu Ile His	
805 810 815	
ggt tgt tcc aaa gat aat aga aat gac aca gaa ggc ttt aag tat cca	2615
Gly Cys Ser Lys Asp Asn Arg Asn Asp Thr Glu Gly Phe Lys Tyr Pro	
820 825 830	
ttg gga cat gaa gtt aac cac agt cgg gaa aca agc ata gaa atg gaa	2663
Leu Gly His Glu Val Asn His Ser Arg Glu Thr Ser Ile Glu Met Glu	
835 840 845	
gaa agt gaa ctt gat gct cag tat ttg cag aat aca ttc aag gtt tca	2711
Glu Ser Glu Leu Asp Ala Gln Tyr Leu Gln Asn Thr Phe Lys Val Ser	
850 855 860	
aag cgc cag tca ttt gct ctg ttt tca aat cca gga aat gca gaa gag	2759
Lys Arg Gln Ser Phe Ala Leu Phe Ser Asn Pro Gly Asn Ala Glu Glu	
865 870 875 880	
gaa tgt gca aca ttc tct gcc cac tct ggg tcc tta aag aaa caa agt	2807
Glu Cys Ala Thr Phe Ser Ala His Ser Gly Ser Leu Lys Lys Gln Ser	
885 890 895	

A000626

US 6,951,721 B2

159

160

-continued

cca aaa gtc act ttt gaa tgt gaa caa aag gaa gaa aat caa gga aag Pro Lys Val Thr Phe Glu Cys Glu Gln Lys Glu Glu Asn Gln Gly Lys 900 905 910	2855
aat gag tct aat atc aag cct gta cag aca gtt aat atc act gca ggc Asn Glu Ser Asn Ile Lys Pro Val Gln Thr Val Asn Ile Thr Ala Gly 915 920 925	2903
ttt cct gtg gtt ggt cag aaa gat aag cca gtt gat aat gcc aaa tgt Phe Pro Val Val Gly Gln Lys Asp Lys Pro Val Asp Asn Ala Lys Cys 930 935 940	2951
agt atc aaa gga ggc tct agg ttt tgt cta tca tct cag ttc aga ggc Ser Ile Lys Gly Gly Ser Arg Phe Cys Leu Ser Ser Gln Phe Arg Gly 945 950 955 960	2999
aac gaa act gga ctc att act cca aat aaa cat gga ctt tta caa aac Asn Glu Thr Gly Leu Ile Thr Pro Asn Lys His Gly Leu Leu Gln Asn 965 970 975	3047
cca tat cgt ata cca cca ctt ttt ccc atc aag tca ttt gtt aaa act Pro Tyr Arg Ile Pro Pro Leu Phe Pro Ile Lys Ser Phe Val Lys Thr 980 985 990	3095
aaa tgt aag aaa aat ctg cta gag gaa aac ttt gag gaa cat tca atg Lys Cys Lys Lys Asn Leu Leu Glu Glu Asn Phe Glu Glu His Ser Met 995 1000 1005	3143
tca cct gaa aga gaa atg gga aat gag aac att cca agt aca gtg Ser Pro Glu Arg Glu Met Gly Asn Glu Asn Ile Pro Ser Thr Val 1010 1015 1020	3188
agc aca att agc cgt aat aac att aga gaa aat gtt ttt aaa gaa Ser Thr Ile Ser Arg Asn Asn Ile Arg Glu Asn Val Phe Lys Glu 1025 1030 1035	3233
gcc agc tca agc aat att aat gaa gta ggt tcc agt act aat gaa Ala Ser Ser Ser Asn Ile Asn Glu Val Gly Ser Ser Thr Asn Glu 1040 1045 1050	3278
gtg ggc tcc agt att aat gaa ata ggt tcc agt gat gaa aac att Val Gly Ser Ser Ile Asn Glu Ile Gly Ser Ser Asp Glu Asn Ile 1055 1060 1065	3323
caa gca gaa cta ggt aga aac aga ggg cca aaa ttg aat gct atg Gln Ala Glu Leu Gly Arg Asn Arg Gly Pro Lys Leu Asn Ala Met 1070 1075 1080	3368
ctt aga tta ggg gtt ttg caa cct gag gtc tat aaa caa agt ctt Leu Arg Leu Gly Val Leu Gln Pro Glu Val Tyr Lys Gln Ser Leu 1085 1090 1095	3413
cct gga agt aat tgt aag cat cct gaa ata aaa aag caa gaa tat Pro Gly Ser Asn Cys Lys His Pro Glu Ile Lys Lys Gln Glu Tyr 1100 1105 1110	3458
gaa gaa gta gtt cag act gtt aat aca gat ttc tct cca tat ctg Glu Glu Val Val Gln Thr Val Asn Thr Asp Phe Ser Pro Tyr Leu 1115 1120 1125	3503
att tca gat aac tta gaa cag cct atg gga agt agt cat gca tct Ile Ser Asp Asn Leu Glu Gln Pro Met Gly Ser Ser His Ala Ser 1130 1135 1140	3548
cag gtt tgt tct gag aca cct gat gac ctg tta gat gat ggt gaa Gln Val Cys Ser Glu Thr Pro Asp Asp Leu Leu Asp Asp Gly Glu 1145 1150 1155	3593
ata aag gaa gat act agt ttt gct gaa aat gac att aag gaa agt Ile Lys Glu Asp Thr Ser Phe Ala Glu Asn Asp Ile Lys Glu Ser 1160 1165 1170	3638
tct gct gtt ttt agc aaa agc gtc cag aaa gga gag ctt agc agg Ser Ala Val Phe Ser Lys Ser Val Gln Lys Gly Glu Leu Ser Arg 1175 1180 1185	3683
agt cct agc cct ttc acc cat aca cat ttg gct cag ggt tac cga Ser Pro Ser Pro Phe Thr His Thr His Leu Ala Gln Gly Tyr Arg 1190 1195 1200	3728

A000627

US 6,951,721 B2

161

162

-continued

aga ggg gcc aag aaa tta gag tcc tca gaa gag aac tta tct agt 3773 Arg Gly Ala Lys Lys Leu Glu Ser Ser Glu Glu Asn Leu Ser Ser 1205 1210 1215
gag gat gaa gag ctt ccc tgc ttc caa cac ttg tta ttt ggt aaa 3818 Glu Asp Glu Glu Leu Pro Cys Phe Gln His Leu Leu Phe Gly Lys 1220 1225 1230
gta aac aat ata cct tct cag tct act agg cat agc acc gtt gct 3863 Val Asn Asn Ile Pro Ser Ser Gln Ser Thr Arg His Ser Thr Val Ala 1235 1240 1245
acc gag tgt ctg tct aag aac aca gag gag aat tta tta tca ttg 3908 Thr Glu Cys Leu Ser Lys Asn Thr Glu Glu Asn Leu Leu Ser Leu 1250 1255 1260
aag aat agc tta aat gac tgc agt aac cag gta ata ttg gca aag 3953 Lys Asn Ser Leu Asn Asp Cys Ser Asn Gln Val Ile Leu Ala Lys 1265 1270 1275
gca tct cag gaa cat cac ctt agt gag gaa aca aaa tgt tct gct 3998 Ala Ser Gln Glu His His Leu Ser Glu Glu Thr Lys Cys Ser Ala 1280 1285 1290
agc ttg ttt tct tca cag tgc agt gaa ttg gaa gac ttg act gca 4043 Ser Leu Phe Ser Ser Gln Cys Ser Glu Leu Glu Asp Leu Thr Ala 1295 1300 1305
aat aca aac acc cag gat cct ttc ttg att ggt tct tcc aaa caa 4088 Asn Thr Asn Thr Gln Asp Pro Phe Leu Ile Gly Ser Ser Lys Gln 1310 1315 1320
atg agg cat cag tct gaa agc cag gga gtt ggt ctg agt gac aag 4133 Met Arg His Gln Ser Glu Ser Gln Gly Val Gly Leu Ser Asp Lys 1325 1330 1335
gaa ttg gtt tca gat gat gaa gaa aga gga acg ggc ttg gaa gaa 4178 Glu Leu Val Ser Asp Asp Glu Glu Arg Gly Thr Gly Leu Glu Glu 1340 1345 1350
aat aat caa gaa gag caa agc atg gat tca aac tta ggt gaa gca 4223 Asn Asn Gln Glu Glu Gln Ser Met Asp Ser Asn Leu Gly Glu Ala 1355 1360 1365
gca tct ggg tgt gag agt gaa aca agc gtc tct gaa gac tgc tca 4268 Ala Ser Gly Cys Glu Ser Glu Thr Ser Val Ser Glu Asp Cys Ser 1370 1375 1380
ggg cta tcc tct cag agt gac att tta acc act cag cag agg gat 4313 Gly Leu Ser Ser Gln Ser Asp Ile Leu Thr Thr Gln Gln Arg Asp 1385 1390 1395
acc atg caa cat aac ctg ata aag ctc cag cag gaa atg gct gaa 4358 Thr Met Gln His Asn Leu Ile Lys Leu Gln Gln Glu Met Ala Glu 1400 1405 1410
cta gaa gct gtg tta gaa cag cat ggg agc cag cct tct aac agc 4403 Leu Glu Ala Val Leu Glu Gln His Gly Ser Gln Pro Ser Asn Ser 1415 1420 1425
tac cct tcc atc ata agt gac tct tct gcc ctt gag gac ctg cga 4448 Tyr Pro Ser Ile Ile Ser Asp Ser Ser Ala Leu Glu Asp Leu Arg 1430 1435 1440
aat cca gaa caa agc aca tca gaa aaa gca gta tta act tca cag 4493 Asn Pro Glu Gln Ser Thr Ser Glu Lys Ala Val Leu Thr Ser Gln 1445 1450 1455
aaa agt agt gaa tac cct ata agc cag aat cca gaa ggc ctt tct 4538 Lys Ser Ser Glu Tyr Pro Ile Ser Gln Asn Pro Glu Gly Leu Ser 1460 1465 1470
gct gac aag ttt gag gtg tct gca gat agt tct acc agt aaa aat 4583 Ala Asp Lys Phe Glu Val Ser Ala Asp Ser Ser Thr Ser Lys Asn 1475 1480 1485
aaa gaa cca gga gtg gaa agg tca tcc cct tct aaa tgc cca tca 4628 Lys Glu Pro Gly Val Glu Arg Ser Ser Pro Ser Lys Cys Pro Ser 1490 1495 1500

A000628

US 6,951,721 B2

163

164

-continued

1490	1495	1500	
tta gat gat agg tgg tac atg cac agt tgc tct ggg agt ctt cag 4673			
Leu Asp Asp Arg Trp Tyr Met His Ser Cys Ser Gly Ser Leu Gln			
1505	1510	1515	
aat aga aac tac cca tct caa gag gag ctc att aag gtt gtt gat 4718			
Asn Arg Asn Tyr Pro Ser Gln Glu Glu Leu Ile Lys Val Val Asp			
1520	1525	1530	
gtg gag gag caa cag ctg gaa gag tct ggg cca cac gat ttg acg 4763			
Val Glu Glu Gln Gln Leu Glu Glu Ser Gly Pro His Asp Leu Thr			
1535	1540	1545	
gaa aca tct tac ttg cca agg caa gat cta gag gga acc cct tac 4808			
Glu Thr Ser Tyr Leu Pro Arg Gln Asp Leu Glu Gly Thr Pro Tyr			
1550	1555	1560	
ctg gaa tct gga atc agc ctc ttc tct gat gac cct gaa tct gat 4853			
Leu Glu Ser Gly Ile Ser Leu Phe Ser Asp Asp Pro Glu Ser Asp			
1565	1570	1575	
cct tct gaa gac aga gcc cca gag tca gct cgt gtt ggc aac ata 4898			
Pro Ser Glu Asp Arg Ala Pro Glu Ser Ala Arg Val Gly Asn Ile			
1580	1585	1590	
cca tct tca acc tct gca ttg aaa gtt ccc caa ttg aaa gtt gca 4943			
Pro Ser Ser Thr Ser Ala Leu Lys Val Pro Gln Leu Lys Val Ala			
1595	1600	1605	
gaa tct gcc cag agt cca gct gct gct cat act act gat act gct 4988			
Glu Ser Ala Gln Ser Pro Ala Ala Ala His Thr Thr Asp Thr Ala			
1610	1615	1620	
ggg tat aat gca atg gaa gaa agt gtg agc agg gag aag cca gaa 5033			
Gly Tyr Asn Ala Met Glu Glu Ser Val Ser Arg Glu Lys Pro Glu			
1625	1630	1635	
ttg aca gct tca aca gaa agg gtc aac aaa aga atg tcc atg gtg 5078			
Leu Thr Ala Ser Thr Glu Arg Val Asn Lys Arg Met Ser Met Val			
1640	1645	1650	
gtg tct ggc ctg acc cca gaa gaa ttt atg ctc gtg tac aag ttt 5123			
Val Ser Gly Leu Thr Pro Glu Glu Phe Met Leu Val Tyr Lys Phe			
1655	1660	1665	
gcc aga aaa cac cac atc act tta act aat cta att act gaa gag 5168			
Ala Arg Lys His His Ile Thr Leu Thr Asn Leu Ile Thr Glu Glu			
1670	1675	1680	
act act cat gtt gtt atg aaa aca gat gct gag ttt gtg tgt gaa 5213			
Thr Thr His Val Val Met Lys Thr Asp Ala Glu Phe Val Cys Glu			
1685	1690	1695	
cgg aca ctg aaa tat ttt cta gga att gcg gga gga aaa tgg gta 5258			
Arg Thr Leu Lys Tyr Phe Leu Gly Ile Ala Gly Gly Lys Trp Val			
1700	1705	1710	
gtt agc tat ttc tgg gtg acc cag tct att aaa gaa aga aaa atg 5303			
Val Ser Tyr Phe Trp Val Thr Gln Ser Ile Lys Glu Arg Lys Met			
1715	1720	1725	
ctg aat gag cat gat ttt gaa gtc aga gga gat gtg gtc aat gga 5348			
Leu Asn Glu His Asp Phe Glu Val Arg Gly Asp Val Val Asn Gly			
1730	1735	1740	
aga aac cac caa ggt cca aag cga gca aga gaa tcc cag gac aga 5393			
Arg Asn His Gln Gly Pro Lys Arg Ala Arg Glu Ser Gln Asp Arg			
1745	1750	1755	
aag atc ttc agg ggg cta gaa atc tgt tgc tat ggg ccc ttc acc 5438			
Lys Ile Phe Arg Gly Leu Glu Ile Cys Cys Tyr Gly Pro Phe Thr			
1760	1765	1770	
aac atg ccc aca gat caa ctg gaa tgg atg gta cag ctg tgt ggt 5483			
Asn Met Pro Thr Asp Gln Leu Glu Trp Met Val Gln Leu Cys Gly			
1775	1780	1785	
gct tct gtg gtg aag gag ctt tca tca ttc acc ctt ggc aca ggt 5528			

A000629

US 6,951,721 B2

165

166

-continued

Ala Ser	Val Val Lys	Glu Leu	Ser Ser Phe Thr	Leu Gly Thr Gly	
1790		1795		1800	
gtc cac	cca att gtg gtt gtg	cag cca gat gcc tgg	aca gag gac	5573	
Val His	Pro Ile Val Val Val	Gln Pro Asp Ala Trp	Thr Glu Asp		
1805		1810	1815		
aat ggc	ttc cat gca att ggg	cag atg tgt gag gca	cct gtg gtg	5618	
Asn Gly	Phe His Ala Ile Gly	Gln Met Cys Glu Ala	Pro Val Val		
1820		1825	1830		
acc cga	gag tgg gtg ttg gac	agt gta gca ctc tac	cag tgc cag	5663	
Thr Arg	Glu Trp Val Leu Asp	Ser Val Ala Leu Tyr	Gln Cys Gln		
1835		1840	1845		
gag ctg	gac acc tac ctg ata	ccc cag atc ccc cac	agc cac tac	5708	
Glu Leu	Asp Thr Tyr Leu Ile	Pro Gln Ile Pro His	Ser His Tyr		
1850		1855	1860		
tga				5711	
<210> SEQ ID NO 268					
<211> LENGTH: 1863					
<212> TYPE: PRT					
<213> ORGANISM: Homo sapiens					
<400> SEQUENCE: 268					
Met Asp Leu Ser	Ala Leu Arg Val	Glu Glu Val Gln	Asn Val Ile Asn		
1	5	10	15		
Ala Met Gln Lys	Ile Leu Glu Cys	Pro Ile Cys Leu	Glu Leu Ile Lys		
	20	25	30		
Glu Pro Val Ser	Thr Lys Cys Asp	His Ile Phe Cys	Lys Phe Cys Met		
	35	40	45		
Leu Lys Leu Leu	Asn Gln Lys Lys	Gly Pro Ser Gln	Cys Pro Leu Cys		
	50	55	60		
Lys Asn Asp Ile	Thr Lys Arg Ser	Leu Gln Glu Ser	Thr Arg Phe Ser		
65	70	75	80		
Gln Leu Val Glu	Glu Leu Leu Lys	Ile Ile Cys Ala	Phe Gln Leu Asp		
	85	90	95		
Thr Gly Leu Glu	Tyr Ala Asn Ser	Tyr Asn Phe Ala	Lys Lys Glu Asn		
	100	105	110		
Asn Ser Pro Glu	His Leu Lys Asp	Glu Val Ser Ile	Ile Gln Ser Met		
	115	120	125		
Gly Tyr Arg Asn	Arg Ala Lys Arg	Leu Leu Gln Ser	Glu Pro Glu Asn		
130	135	140			
Pro Ser Leu Gln	Glu Thr Ser Leu	Ser Val Gln Leu	Ser Asn Leu Gly		
145	150	155	160		
Thr Val Arg Thr	Leu Arg Thr Lys	Gln Arg Ile Gln	Pro Gln Lys Thr		
	165	170	175		
Ser Val Tyr Ile	Glu Leu Gly Ser	Asp Ser Ser Glu	Asp Thr Val Asn		
	180	185	190		
Lys Ala Thr Tyr	Cys Ser Val Gly	Asp Gln Glu Leu	Leu Gln Ile Thr		
	195	200	205		
Pro Gln Gly Thr	Arg Asp Glu Ile	Ser Leu Asp Ser	Ala Lys Lys Ala		
	210	215	220		
Ala Cys Glu Phe	Ser Glu Thr Asp	Val Thr Asn Thr	Glu His His Gln		
225	230	235	240		
Pro Ser Asn Asn	Asp Leu Asn Thr	Thr Glu Lys Arg	Ala Ala Glu Arg		
	245	250	255		
His Pro Glu Lys	Tyr Gln Gly Ser	Ser Val Ser Asn	Leu His Val Glu		
	260	265	270		

A000630

US 6,951,721 B2

167

168

-continued

Pro Cys Gly Thr Asn Thr His Ala Ser Ser Leu Gln His Glu Asn Ser
 275 280 285
 Ser Leu Leu Leu Thr Lys Asp Arg Met Asn Val Glu Lys Ala Glu Phe
 290 295 300
 Cys Asn Lys Ser Lys Gln Pro Gly Leu Ala Arg Ser Gln His Asn Arg
 305 310 315 320
 Trp Ala Gly Ser Lys Glu Thr Cys Asn Asp Arg Arg Thr Pro Ser Thr
 325 330 335
 Glu Lys Lys Val Asp Leu Asn Ala Asp Pro Leu Cys Glu Arg Lys Glu
 340 345 350
 Trp Asn Lys Gln Lys Leu Pro Cys Ser Glu Asn Pro Arg Asp Thr Glu
 355 360 365
 Asp Val Pro Trp Ile Thr Leu Asn Ser Ser Ile Gln Lys Val Asn Glu
 370 375 380
 Trp Phe Ser Arg Ser Asp Glu Leu Leu Gly Ser Asp Asp Ser His Asp
 385 390 395 400
 Gly Glu Ser Glu Ser Asn Ala Lys Val Ala Asp Val Leu Asp Val Leu
 405 410 415
 Asn Glu Val Asp Glu Tyr Ser Gly Ser Ser Glu Lys Ile Asp Leu Leu
 420 425 430
 Ala Ser Asp Pro His Glu Ala Leu Ile Cys Lys Ser Glu Arg Val His
 435 440 445
 Ser Lys Ser Val Glu Ser Asn Ile Glu Asp Lys Ile Phe Gly Lys Thr
 450 455 460
 Tyr Arg Lys Lys Ala Ser Leu Pro Asn Leu Ser His Val Thr Glu Asn
 465 470 475 480
 Leu Ile Ile Gly Ala Phe Val Thr Glu Pro Gln Ile Ile Gln Glu Arg
 485 490 495
 Pro Leu Thr Asn Lys Leu Lys Arg Lys Arg Arg Pro Thr Ser Gly Leu
 500 505 510
 His Pro Glu Asp Phe Ile Lys Lys Ala Asp Leu Ala Val Gln Lys Thr
 515 520 525
 Pro Glu Met Ile Asn Gln Gly Thr Asn Gln Thr Glu Gln Asn Gly Gln
 530 535 540
 Val Met Asn Ile Thr Asn Ser Gly His Glu Asn Lys Thr Lys Gly Asp
 545 550 555 560
 Ser Ile Gln Asn Glu Lys Asn Pro Asn Pro Ile Glu Ser Leu Glu Lys
 565 570 575
 Glu Ser Ala Phe Lys Thr Lys Ala Glu Pro Ile Ser Ser Ser Ile Ser
 580 585 590
 Asn Met Glu Leu Glu Leu Asn Ile His Asn Ser Lys Ala Pro Lys Lys
 595 600 605
 Asn Arg Leu Arg Arg Lys Ser Ser Thr Arg His Ile His Ala Leu Glu
 610 615 620
 Leu Val Val Ser Arg Asn Leu Ser Pro Pro Asn Cys Thr Glu Leu Gln
 625 630 635 640
 Ile Asp Ser Cys Ser Ser Ser Glu Glu Ile Lys Lys Lys Lys Tyr Asn
 645 650 655
 Gln Met Pro Val Arg His Ser Arg Asn Leu Gln Leu Met Glu Gly Lys
 660 665 670
 Glu Pro Ala Thr Gly Ala Lys Lys Ser Asn Lys Pro Asn Glu Gln Thr
 675 680 685

A000631

US 6,951,721 B2

169

170

-continued

Ser	Lys	Arg	His	Asp	Ser	Asp	Thr	Phe	Pro	Glu	Leu	Lys	Leu	Thr	Asn
690						695					700				
Ala	Pro	Gly	Ser	Phe	Thr	Lys	Cys	Ser	Asn	Thr	Ser	Glu	Leu	Lys	Glu
705					710					715					720
Phe	Val	Asn	Pro	Ser	Leu	Pro	Arg	Glu	Glu	Lys	Glu	Glu	Lys	Leu	Glu
				725				730						735	
Thr	Val	Lys	Val	Ser	Asn	Asn	Ala	Glu	Asp	Pro	Lys	Asp	Leu	Met	Leu
			740					745					750		
Ser	Gly	Glu	Arg	Val	Leu	Gln	Thr	Glu	Arg	Ser	Val	Glu	Ser	Ser	Ser
	755						760					765			
Ile	Ser	Leu	Val	Pro	Gly	Thr	Asp	Tyr	Gly	Thr	Gln	Glu	Ser	Ile	Ser
	770					775					780				
Leu	Leu	Glu	Val	Ser	Thr	Leu	Gly	Lys	Ala	Lys	Thr	Glu	Pro	Asn	Lys
785					790					795					800
Cys	Val	Ser	Gln	Cys	Ala	Ala	Phe	Glu	Asn	Pro	Lys	Gly	Leu	Ile	His
				805					810					815	
Gly	Cys	Ser	Lys	Asp	Asn	Arg	Asn	Asp	Thr	Glu	Gly	Phe	Lys	Tyr	Pro
			820					825					830		
Leu	Gly	His	Glu	Val	Asn	His	Ser	Arg	Glu	Thr	Ser	Ile	Glu	Met	Glu
	835						840					845			
Glu	Ser	Glu	Leu	Asp	Ala	Gln	Tyr	Leu	Gln	Asn	Thr	Phe	Lys	Val	Ser
	850					855					860				
Lys	Arg	Gln	Ser	Phe	Ala	Leu	Phe	Ser	Asn	Pro	Gly	Asn	Ala	Glu	Glu
865					870					875					880
Glu	Cys	Ala	Thr	Phe	Ser	Ala	His	Ser	Gly	Ser	Leu	Lys	Lys	Gln	Ser
				885					890					895	
Pro	Lys	Val	Thr	Phe	Glu	Cys	Glu	Gln	Lys	Glu	Glu	Asn	Gln	Gly	Lys
		900						905					910		
Asn	Glu	Ser	Asn	Ile	Lys	Pro	Val	Gln	Thr	Val	Asn	Ile	Thr	Ala	Gly
	915						920					925			
Phe	Pro	Val	Val	Gly	Gln	Lys	Asp	Lys	Pro	Val	Asp	Asn	Ala	Lys	Cys
	930					935					940				
Ser	Ile	Lys	Gly	Gly	Ser	Arg	Phe	Cys	Leu	Ser	Ser	Gln	Phe	Arg	Gly
945					950					955					960
Asn	Glu	Thr	Gly	Leu	Ile	Thr	Pro	Asn	Lys	His	Gly	Leu	Leu	Gln	Asn
				965					970					975	
Pro	Tyr	Arg	Ile	Pro	Pro	Leu	Phe	Pro	Ile	Lys	Ser	Phe	Val	Lys	Thr
		980						985					990		
Lys	Cys	Lys	Lys	Asn	Leu	Leu	Glu	Glu	Asn	Phe	Glu	Glu	His	Ser	Met
		995					1000					1005			
Ser	Pro	Glu	Arg	Glu	Met	Gly	Asn	Glu	Asn	Ile	Pro	Ser	Thr	Val	
	1010					1015					1020				
Ser	Thr	Ile	Ser	Arg	Asn	Asn	Ile	Arg	Glu	Asn	Val	Phe	Lys	Glu	
	1025					1030					1035				
Ala	Ser	Ser	Ser	Asn	Ile	Asn	Glu	Val	Gly	Ser	Ser	Thr	Asn	Glu	
	1040					1045					1050				
Val	Gly	Ser	Ser	Ile	Asn	Glu	Ile	Gly	Ser	Ser	Asp	Glu	Asn	Ile	
	1055					1060					1065				
Gln	Ala	Glu	Leu	Gly	Arg	Asn	Arg	Gly	Pro	Lys	Leu	Asn	Ala	Met	
	1070					1075					1080				
Leu	Arg	Leu	Gly	Val	Leu	Gln	Pro	Glu	Val	Tyr	Lys	Gln	Ser	Leu	
	1085					1090					1095				
Pro	Gly	Ser	Asn	Cys	Lys	His	Pro	Glu	Ile	Lys	Lys	Gln	Glu	Tyr	

A000632

US 6,951,721 B2

171

172

-continued

1100	1105	1110
Glu Glu Val Val Gln Thr Val Asn Thr Asp Phe Ser Pro Tyr Leu		
1115	1120	1125
Ile Ser Asp Asn Leu Glu Gln Pro Met Gly Ser Ser His Ala Ser		
1130	1135	1140
Gln Val Cys Ser Glu Thr Pro Asp Asp Leu Leu Asp Asp Gly Glu		
1145	1150	1155
Ile Lys Glu Asp Thr Ser Phe Ala Glu Asn Asp Ile Lys Glu Ser		
1160	1165	1170
Ser Ala Val Phe Ser Lys Ser Val Gln Lys Gly Glu Leu Ser Arg		
1175	1180	1185
Ser Pro Ser Pro Phe Thr His Thr His Leu Ala Gln Gly Tyr Arg		
1190	1195	1200
Arg Gly Ala Lys Lys Leu Glu Ser Ser Glu Glu Asn Leu Ser Ser		
1205	1210	1215
Glu Asp Glu Glu Leu Pro Cys Phe Gln His Leu Leu Phe Gly Lys		
1220	1225	1230
Val Asn Asn Ile Pro Ser Gln Ser Thr Arg His Ser Thr Val Ala		
1235	1240	1245
Thr Glu Cys Leu Ser Lys Asn Thr Glu Glu Asn Leu Leu Ser Leu		
1250	1255	1260
Lys Asn Ser Leu Asn Asp Cys Ser Asn Gln Val Ile Leu Ala Lys		
1265	1270	1275
Ala Ser Gln Glu His His Leu Ser Glu Glu Thr Lys Cys Ser Ala		
1280	1285	1290
Ser Leu Phe Ser Ser Gln Cys Ser Glu Leu Glu Asp Leu Thr Ala		
1295	1300	1305
Asn Thr Asn Thr Gln Asp Pro Phe Leu Ile Gly Ser Ser Lys Gln		
1310	1315	1320
Met Arg His Gln Ser Glu Ser Gln Gly Val Gly Leu Ser Asp Lys		
1325	1330	1335
Glu Leu Val Ser Asp Asp Glu Glu Arg Gly Thr Gly Leu Glu Glu		
1340	1345	1350
Asn Asn Gln Glu Glu Gln Ser Met Asp Ser Asn Leu Gly Glu Ala		
1355	1360	1365
Ala Ser Gly Cys Glu Ser Glu Thr Ser Val Ser Glu Asp Cys Ser		
1370	1375	1380
Gly Leu Ser Ser Gln Ser Asp Ile Leu Thr Thr Gln Gln Arg Asp		
1385	1390	1395
Thr Met Gln His Asn Leu Ile Lys Leu Gln Gln Glu Met Ala Glu		
1400	1405	1410
Leu Glu Ala Val Leu Glu Gln His Gly Ser Gln Pro Ser Asn Ser		
1415	1420	1425
Tyr Pro Ser Ile Ile Ser Asp Ser Ser Ala Leu Glu Asp Leu Arg		
1430	1435	1440
Asn Pro Glu Gln Ser Thr Ser Glu Lys Ala Val Leu Thr Ser Gln		
1445	1450	1455
Lys Ser Ser Glu Tyr Pro Ile Ser Gln Asn Pro Glu Gly Leu Ser		
1460	1465	1470
Ala Asp Lys Phe Glu Val Ser Ala Asp Ser Ser Thr Ser Lys Asn		
1475	1480	1485
Lys Glu Pro Gly Val Glu Arg Ser Ser Pro Ser Lys Cys Pro Ser		
1490	1495	1500

A000633

US 6,951,721 B2

173

174

-continued

Leu Asp	Asp Arg	Trp Tyr	Met	His Ser	Cys Ser	Gly	Ser Leu	Gln	
1505			1510			1515			
Asn Arg	Asn Tyr	Pro Ser	Gln	Glu Glu	Leu Ile	Lys	Val Val	Asp	
1520			1525			1530			
Val Glu	Glu Gln	Gln Leu	Glu	Glu Ser	Gly Pro	His	Asp Leu	Thr	
1535			1540			1545			
Glu Thr	Ser Tyr	Leu Pro	Arg	Gln Asp	Leu Glu	Gly	Thr Pro	Tyr	
1550			1555			1560			
Leu Glu	Ser Gly	Ile Ser	Leu	Phe Ser	Asp Asp	Pro	Glu Ser	Asp	
1565			1570			1575			
Pro Ser	Glu Asp	Arg Ala	Pro	Glu Ser	Ala Arg	Val	Gly Asn	Ile	
1580			1585			1590			
Pro Ser	Ser Thr	Ser Ala	Leu	Lys Val	Pro Gln	Leu	Lys Val	Ala	
1595			1600			1605			
Glu Ser	Ala Gln	Ser Pro	Ala	Ala Ala	His Thr	Thr	Asp Thr	Ala	
1610			1615			1620			
Gly Tyr	Asn Ala	Met Glu	Glu	Ser Val	Ser Arg	Glu	Lys Pro	Glu	
1625			1630			1635			
Leu Thr	Ala Ser	Thr Glu	Arg	Val Asn	Lys Arg	Met	Ser Met	Val	
1640			1645			1650			
Val Ser	Gly Leu	Thr Pro	Glu	Glu Phe	Met Leu	Val	Tyr Lys	Phe	
1655			1660			1665			
Ala Arg	Lys His	His Ile	Thr	Leu Thr	Asn Leu	Ile	Thr Glu	Glu	
1670			1675			1680			
Thr Thr	His Val	Val Met	Lys	Thr Asp	Ala Glu	Phe	Val Cys	Glu	
1685			1690			1695			
Arg Thr	Leu Lys	Tyr Phe	Leu	Gly Ile	Ala Gly	Gly	Lys Trp	Val	
1700			1705			1710			
Val Ser	Tyr Phe	Trp Val	Thr	Gln Ser	Ile Lys	Glu	Arg Lys	Met	
1715			1720			1725			
Leu Asn	Glu His	Asp Phe	Glu	Val Arg	Gly Asp	Val	Val Asn	Gly	
1730			1735			1740			
Arg Asn	His Gln	Gly Pro	Lys	Arg Ala	Arg Glu	Ser	Gln Asp	Arg	
1745			1750			1755			
Lys Ile	Phe Arg	Gly Leu	Glu	Ile Cys	Cys Tyr	Gly	Pro Phe	Thr	
1760			1765			1770			
Asn Met	Pro Thr	Asp Gln	Leu	Glu Trp	Met Val	Gln	Leu Cys	Gly	
1775			1780			1785			
Ala Ser	Val Val	Lys Glu	Leu	Ser Ser	Phe Thr	Leu	Gly Thr	Gly	
1790			1795			1800			
Val His	Pro Ile	Val Val	Val	Gln Pro	Asp Ala	Trp	Thr Glu	Asp	
1805			1810			1815			
Asn Gly	Phe His	Ala Ile	Gly	Gln Met	Cys Glu	Ala	Pro Val	Val	
1820			1825			1830			
Thr Arg	Glu Trp	Val Leu	Asp	Ser Val	Ala Leu	Tyr	Gln Cys	Gln	
1835			1840			1845			
Glu Leu	Asp Thr	Tyr Leu	Ile	Pro Gln	Ile Pro	His	Ser His	Tyr	
1850			1855			1860			

A000634

US 6,951,721 B2

175

What is claimed is:

1. A method for determining an omi haplotype of a human BRCA1 gene comprising:

- (a) determining the nucleotide sequence of the BRCA1 gene or fragment thereof from at least one female individual with a family history which indicates a predisposition to breast cancer,
- (b) comparing the determined nucleotide sequence from said female individual to SEQ ID NO: 263, and
- (c) determining the presence of the following nucleotide variations: thymine at nucleotides 2201 and 2731, cytosine at nucleotides 2430 and 4427, and guanine at nucleotides 3232, 3667 and 4956, wherein the presence of the nucleotide variations in the determined nucleotide sequence indicates the omi1 haplotype.

2. The method of claim 1 further comprising repeating steps (a) and (b).

3. The method of claim 1 wherein at least one nucleotide variation is located in an exon coding region of the BRCA1 gene.

4. The method of claim 3 wherein at least one nucleotide variation encodes an amino acid variation in the protein encoded by the BRCA 1 gene.

5. The method of claim 1 wherein the BRCA1 gene or fragment thereof is amplified prior to nucleotide sequencing.

6. The method of claim 1 further comprising comparing the determined nucleotide sequence to SEQ ID NO: 265.

7. The method of claim 1 further comprising comparing the determined nucleotide sequence to SEQ ID NO: 267.

8. The method of claim 1 further comprising determining the putative amino acid sequence of the protein encoded by the BRCA1 gene.

9. The method of claim 8 further comprising comparing the determined putative amino acid sequence to SEQ ID NO: 264.

10. The method of claim 8 further comprising comparing the determined putative amino acid sequence to SEQ ID NO: 266.

11. The method of claim 8 further comprising comparing the determined putative amino acid sequence to SEQ ID NO: 268.

12. The method of claim 1 wherein the nucleotide sequence or fragment thereof of the BRCA1 gene is determined in at least five female individuals with a family history which indicates a predisposition to breast cancer.

13. A method for determining an omi haplotype of a human BRCA1 gene comprising:

- (a) determining the nucleotide sequence of the BRCA1 gene or fragment thereof from at least one female individual with a family history which indicates a predisposition to breast cancer,
- (b) determining the putative amino acid sequence of the protein or fragment thereof encoded by the BRCA1 gene from the determined nucleotide sequence,
- (c) comparing the putative amino acid sequence from said human to SEQ ID NO: 264, and

176

- (d) determining the presence of the following amino acid variations: proline at position 871, glutamate at residue 1038, lysine at residue 1183 and serine at residue 1613 wherein the presence of the variations in the determined amino acid sequence indicates the presence of the omi1 haplotype.

14. The method of claim 13 further comprising comparing the determined putative amino acid sequence to SEQ ID NO: 266.

15. The method of claim 13 further comprising comparing the determined putative amino acid sequence to SEQ ID NO: 268.

16. The method of claim 13 wherein the putative amino acid sequence or fragment thereof of the protein encoded by the BRCA1 gene is determined in at least five female individuals with a family history which indicates a predisposition to breast cancer.

17. A method for determining an omi haplotype of a human BRCA1 gene consisting essentially of:

- (a) determining the nucleotide sequence of the BRCA1 gene or fragment thereof from at least one female individual with a family history which indicates a predisposition to breast cancer,

- (b) comparing the determined nucleotide sequence from said female individual to SEQ ID NO: 263, and

- (c) determining the presence of the following nucleotide variations: thymine at nucleotides 2201 and 2731, cytosine at nucleotides 2430 and 4427, and guanine at nucleotides 3232, 3667 and 4956, wherein the presence of the nucleotide variations in the determined nucleotide sequence indicates the omi1 haplotype.

18. A method for determining an omi haplotype of a human BRCA1 gene consisting essentially of:

- (a) determining the nucleotide sequence of the BRCA1 gene or fragment thereof from at least one female individual with a family history which indicates a predisposition to breast cancer,

- (b) determining the putative amino acid sequence of the protein or fragment thereof encoded by the BRCA1 gene from the determined nucleotide sequence,

- (c) comparing the putative amino acid sequence from said human to SEQ ID NO: 264, and

- (d) determining the presence of the following amino acid variations: proline at position 871, glutamate at residue 1038, lysine at residue 1183 and serine at residue 1613 wherein the presence of the variations in the determined amino acid sequence indicates the presence of the omi1 haplotype.

19. The method according to any of claim 1, 13, 17 or 18 wherein the determined omi1 haplotype of the human BRCA1 gene is not associated with a predisposition to developing breast cancer.

* * * * *

A000635

CERTIFICATE OF SERVICE

I hereby certify that I electronically filed the foregoing with the Clerk of the Court for the United States Court of Appeals for the Federal Circuit by using the appellate CM/ECF system on April 18, 2014:

APPELLANTS' OPENING BRIEF

I further certify that counsel of record are registered as CM/ECF users and will be served by the appellate CM/ECF system and via email:

/s/ Jonathan E. Singer

Jonathan E. Singer

Fish & Richardson P.C.

60 South Sixth Street, Suite 3200

Minneapolis, MN 55402

Tel: (612) 335-5070

Fax: (612) 288-9696

COUNSEL FOR APPELLANT

MYRIAD GENETICS, INC.

CERTIFICATE OF COMPLIANCE

The Opening Brief for Appellant Myriad Genetics, Inc. complies with the type-volume limitation set forth in FRAP 32(a)(7)B). The relevant portions of Appellant's Opening Brief, including all footnotes, contain 13,954 words, as determined by Microsoft Word® 2010.

/s/ Jonathan E. Singer

Jonathan E. Singer
Fish & Richardson P.C.
60 South Sixth Street, Suite 3200
Minneapolis, MN 55402
Tel: (612) 335-5070
Fax: (612) 288-9696

COUNSEL FOR APPELLANT
MYRIAD GENETICS, INC.